



The Effect of Interocular Contrast Differences on the Appearance of Augmented Reality Imagery

MINQI WANG, JIAN DING, DENNIS M. LEVI, and EMILY A. COOPER, University of California, Berkeley

Augmented reality (AR) devices seek to create compelling visual experiences that merge virtual imagery with the natural world. These devices often rely on wearable near-eye display systems that can optically overlay digital images to the left and right eyes of the user separately. Ideally, the two eyes should be shown images with minimal radiometric differences (e.g., the same overall luminance, contrast, and color in both eyes), but achieving this binocular equality can be challenging in wearable systems with stringent demands on weight and size. Basic vision research has shown that a spectrum of potentially detrimental perceptual effects can be elicited by imagery with radiometric differences between the eyes, but it is not clear whether and how these findings apply to the experience of modern AR devices. In this work, we first develop a testing paradigm for assessing multiple aspects of visual appearance at once, and characterize five key perceptual factors when participants viewed stimuli with interocular contrast differences. In a second experiment, we simulate optical see-through AR imagery using conventional desktop LCD monitors and use the same paradigm to evaluate the multi-faceted perceptual implications when the AR display luminance differs between the two eyes. We also include simulations of monocular AR systems (i.e., systems in which only one eye sees the displayed image). Our results suggest that interocular contrast differences can drive several potentially detrimental perceptual effects in binocular AR systems, such as binocular luster, rivalry, and spurious depth differences. In addition, monocular AR displays tend to have more artifacts than binocular displays with a large contrast difference in the two eyes. A better understanding of the range and likelihood of these perceptual phenomena can help inform design choices that support high-quality user experiences in AR.

CCS Concepts: • **Computing methodologies** → **Perception**; Mixed/augmented reality;

Additional Key Words and Phrases: Stereoscopic displays, binocular vision

ACM Reference format:

Minqi Wang, Jian Ding, Dennis M. Levi, and Emily A. Cooper. 2023. The Effect of Interocular Contrast Differences on the Appearance of Augmented Reality Imagery. *ACM Trans. Appl. Percept.* 21, 1, Article 1 (December 2023), 23 pages. <https://doi.org/10.1145/3617684>

1 INTRODUCTION

Designing new display systems often requires understanding whether and how the display's visual limitations adversely affect the user experience. Display systems for **augmented reality (AR)** pose a unique set of challenges because they aim to merge virtual information into the user's natural vision using a system with demanding

This work was supported by a grant from the National Science Foundation (Award 2041726), an NIH training grant (T32EY007043), and a fellowship from the Center for Innovation in Vision and Optics at University of California, Berkeley.

Authors' address: M. Wang, J. Ding, D. M. Levi, and E. A. Cooper, Herbert Wertheim School of Optometry & Vision Science, University of California, Berkeley, Berkeley, CA 94720-2020; e-mails: mwang67@berkeley.edu, jian.ding@berkeley.edu, dlevi@berkeley.edu, emilycooper@berkeley.edu.



This work is licensed under a Creative Commons Attribution International 4.0 License.

© 2023 Copyright held by the owner/author(s).

1544-3558/2023/12-ART1

<https://doi.org/10.1145/3617684>

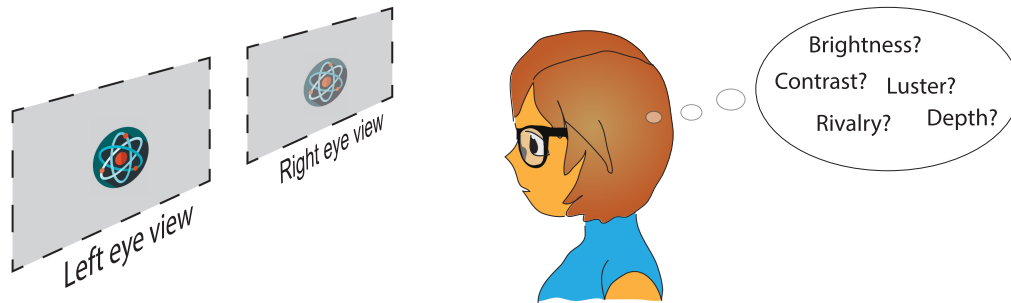


Fig. 1. Binocular display systems present separate images to each eye. These systems are commonly used for augmented reality (AR) and, due to hardware and software limitations or imperfections, are subject to unintended spatial, temporal, or radiometric differences between the images shown to the two eyes. In this illustration, the user's view of an icon (image credit [48]) has higher contrast in the left eye than in the right eye. These differences may affect their perception of brightness, contrast, luster, rivalry, and depth.

design specifications (e.g., a wearable optical see-through near-eye display system) [27]. When these wearable systems are binocular, they may employ independent displays and optics for the two eyes, introducing the potential for spatial, temporal, and radiometric differences in the virtual content that each eye sees (Figure 1). Here, we aim to explore the range of perceptual effects that can result when the user of an AR system receives a higher intensity image in one eye than the other.

From a display engineering perspective, differences between the left and right eye's views can be desirable or detrimental. Importantly, binocular display systems enable the presentation of images with binocular disparities—the natural spatial offsets between the two eyes' views that can elicit a compelling sense of depth via a perceptual process called stereopsis. However, patterns of imperfections in display panels (sometimes called mura) and spatial distortions introduced by optical architectures may also differ between the two eyes [26, 28, 38, 50]. These factors can introduce additional interocular differences that are not intended by the designer, and understanding their potential perceptual consequences is key for optimizing the user experience. Our understanding of these perceptual consequences, however, is still in the early stages.

Basic vision science studies, using simple shapes and gratings as stimuli, suggest that large interocular differences in brightness, contrast, and pattern between the two eyes are likely to elicit troublesome percepts in which the stimulus appears to shimmer or alternate in appearance over time (see [3, 54] for review articles). However, small differences can go unnoticed [14]. Recent applied research has begun exploring whether and how these phenomena might affect the appearance of AR content. For example, in AR systems with a small eyepiece, the two eyes can be subject to different patterns of luminance vignetting (non-uniformity), which may result in degraded image quality [4]. However, a recent perceptual study suggests that a binocular AR display system with different vignetting patterns between the two eyes results in reduced salience of these artifacts as compared to a monocular system [8]. This prior work shows that certain types of radiometric differences may not be detrimental; in fact, it may be possible to take advantage of binocular combination to achieve certain desired design goals (i.e., better display uniformity). However, prior work has focused largely on assessing just one aspect of binocular perceptual experience at a time [2, 9, 10, 20, 29, 53], while it is likely that binocular image differences cause multi-faceted perceptual effects that are not well captured by a single perceptual measurement.

Here, we adopt the term *dichoptic* to refer to stimuli that differ radiometrically between the two eyes (e.g., differ in luminance, contrast, or color). We aim to contribute a better understanding of the perceptual phenomena that occur when viewing dichoptic imagery in AR, in order to support well-informed display design decisions.

We conducted two perceptual experiments to evaluate the implications of dichoptic imagery for user experience in AR systems using a desktop monitor setup that simulates AR imagery. In addition to exploring dichoptic

contrast in binocular AR systems, we include conditions simulating monocular AR viewing (i.e., systems in which only one eye sees the displayed image), because monocular designs may be sufficient for some AR applications. Drawing on the basic vision science literature, we identified several perceptual factors pertinent to the appearance of dichoptic imagery: perceived brightness, contrast, luster, rivalry, and depth (Figure 1). Using simulated AR stimuli, we studied all the aforementioned effects together with a battery of subjective response prompts. We examined how these effects varied for different levels of interocular difference and different stimulus patterns. Studying these factors all together, rather than focusing on a single effect, enables us to characterize a broad gamut of potential perceptual consequences to AR display design.

While many of the presented results are relevant to **virtual reality (VR)** as well, we focus on optical see-through AR in this report because the optical and electronic demands on such systems necessitate challenging tradeoffs that can be informed by a deeper understanding of dichoptic perception. For example, because AR devices often have light pass through from the environment while VR devices block visible light from the environment, AR devices may need higher intensity light sources for content to be visible when operating in bright environments (video see-through AR devices are an exception). AR imagery presented on optical see-through devices also has an idiosyncratic appearance because the optically overlaid virtual content is often semitransparent. Stimuli with this appearance warrant dedicated investigation as our perceptual interpretations of them are complex [17, 57] and underexplored.

Our primary findings are as follows:

- (1) Across a broad range of visual stimuli, participants judged the appearance of dichoptic images to differ from non-dichoptic images with respect to all five perceptual factors tested. We found that luster was the perceptual effect reported most often with dichoptic stimuli.
- (2) As the contrast difference between the two eyes increased, the prevalence of all dichoptic perceptual effects increased.
- (3) Monocular viewing (i.e., viewing display content in only one eye) resulted in a similar set of perceptual results, but with a higher prevalence.

2 RELATED WORK

In this section, we briefly summarize the range of perceptual phenomena associated with viewing dichoptic stimuli that differ in luminance or contrast between the two eyes. We focus on achromatic imagery, but these effects are also relevant for chromatic stimuli, which we will take up briefly in the Discussion.

2.1 Brightness

For most people, closing one eye does not make the world appear any dimmer under normal viewing conditions. This observation suggests that perceived brightness is not a simple average of the luminance levels reaching the two eyes and has motivated a range of psychophysical research characterizing dichoptic brightness perception. Brightness perception is well-modeled as a weighted combination of the inputs to the two eyes, with the weights varying depending on the context. For example, when simple stimuli (e.g., uniform gray disks) with different luminance levels are shown to the two eyes, we can ask what the resulting perceived brightness is. Generally, the stimulus with a greater contrast is found to dominate the binocular brightness percept (sometimes termed “winner-take-all”) [2, 10, 30] (Figure 2(a)). That is, if both stimuli are bright compared to the background (increments), the binocularly perceived brightness tends to match the brighter stimulus and if both stimuli are dark compared to the background (decrements), the perceived brightness matches the darker one. However, percepts can shift toward binocular averaging under certain viewing situations [10]. For example, in *Fechner’s paradox*, viewing a dichoptic image pair with different luminance levels in the two eyes results in a darker percept than if the observer closes one eye and just views the brighter of the pair monocularly [30]. Under certain viewing conditions, the brightness percept can be more like “loser-take-all” and biased toward the dimmer image. In particular, if additional contours or edges are added to the stimulus with lower contrast, the perceptual biases can

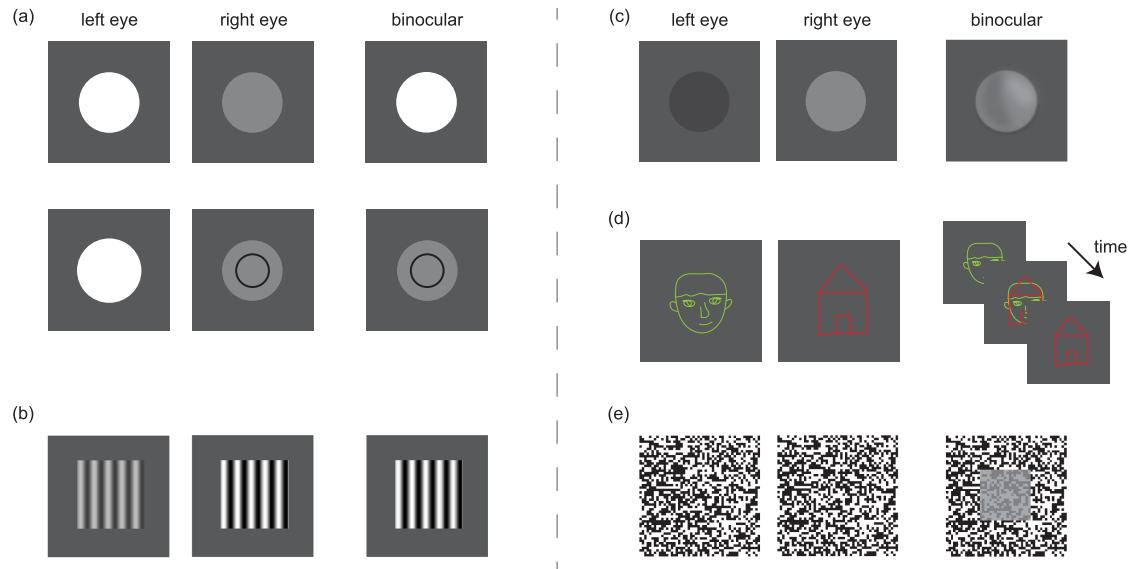


Fig. 2. Illustrations of the different binocular perceptual phenomena that can result from interocular luminance and contrast differences. Readers are encouraged to cross-fuse the left and right eye images to observe the effects since the artistic depiction is not exact. (a) Two pairs of stimuli with dichoptic luminance increments. The top row shows the winner-take-all brightness perception phenomenon. The bottom row, with a monocular contour in the eye seeing the lower luminance disk, illustrates the resulting bias toward that eye (in this case, loser-take-all). (b) Dichoptic contrast perception of more complex patterns, for example if the contrast of a grating pattern differs between the two eyes, is often dominated by the eye seeing higher contrast. (c) Binocular luster percepts can be elicited by dichoptic luminance stimuli. (d) Binocular rivalry can be elicited by pairs of images with different visual patterns in the two eyes. (e) Lastly, imagery that is anti-correlated and without binocular disparity between the two eyes can result in anomalous depth percepts. In this example, the binocular percept of the middle region (highlighted by the gray square) tends to be that it is at a different depth than the surrounding area. Luster, rivalry, and anomalous depth may also be visible when fusing panels (a) and (b).

switch toward that stimulus [10, 30]. These observations motivate the need to understand brightness perception in AR systems. For example, if Fechner’s paradox or loser-take-all binocular combination occurs for AR displays, then it may be better in some cases to have a monocular AR display system than a binocular one with dichoptic brightness.

2.2 Contrast

A related line of research has asked how people perceive the contrast of dichoptic stimuli when the average luminance is matched between the two eyes. Contrast refers to the range between the brightest and darkest regions of an image. For example, research participants can be asked to match or rate the perceived contrast of a binocularly viewed sine wave grating when the two eyes view gratings with different contrast levels (Figure 2(b)). Research using this type of stimulus has shown that dichoptic contrast perception also tends to follow a winner-take-all pattern similar to dichoptic luminance perception [9, 29]. This finding is not surprising given that a sine wave grating can be thought of as a set of alternating luminance increments and decrements. Even when the phase between the two dichoptic gratings differs, the perceived contrast is still biased toward the higher contrast grating [9, 23]. From a display design perspective, these findings seem promising because they suggest that winner-take-all binocular contrast perception can hold even for stereoscopic stimuli with binocular disparities. However, like luminance perception, contextual effects can alter the balance between the two eyes for perceived

contrast. For example, recent work showed that the dichoptic contrast percept can be strongly influenced by a lower contrast stimulus if it is embedded within a contour, similar to brightness percepts [52]. These modulations were also found to depend on the spatial properties of the stimulus: the influence of the contour was stronger for simple grating-like stimuli with a single orientation and weaker for other more complex stimuli.

2.3 Luster

If our perception of dichoptic stimuli could be completely modeled as a weighted mixture of the luminance and contrast of the two eyes inputs, then the challenge of predicting these percepts would just be a matter of determining the appropriate weights for a given stimulus. However, this is not the case. There are unique forms of binocular appearance that can emerge with dichoptic stimuli, such as binocular luster. The *lustrous* appearance of dichoptic stimuli is subjectively described as shimmery, shiny, or metallic (Figure 2(c)). A classic stimulus targeted to elicit binocular luster is a pattern with opposite contrast polarity in the two eyes (i.e., a luminance increment in one eye and decrement in the other eye), but binocular luster can also be elicited when the two eyes have unequal increments or decrements [36, 53]. For AR applications, luster may be troublesome if it interferes with the perceived realism or material properties of the stimulus. However, it may also be a tool that designers desire to leverage, for example, to make a virtual object stand out visually or to break color metamerism [16] (see [54] for review).

2.4 Binocular Rivalry

Another binocular phenomenon that occurs with dichoptic imagery is rivalry. During binocular rivalry, the appearance of a stimulus changes over time. When a stimulus elicits binocular rivalry, it may appear to match one or the other eye's input at any moment in time, or it may be perceived as a mixture (Figure 2(d)). For example, the binocular percept may be a patchy mix of the two eyes inputs, in which some parts of the percept look like one eye's input while other parts look like the other eye's input [46]. To study binocular rivalry, a pair of highly dissimilar images (e.g., gratings with different orientations or two disparate images) are often used, but rivalry can also be elicited by more subtle interocular differences [42]. For many binocular AR devices, it is unlikely that the content seen by the two eyes is extremely dissimilar, but for monocular devices that show virtual content to only one eye rivalry may be more of a concern [40]. It is thought that the relative strength of each eye's input determines rivalry dynamics, and the eye with the stronger stimulus (e.g., brighter, higher contrast) is the predominant percept [31]. This observation holds true for simple stimuli, but not necessarily for more complex stimuli [47]. Compared to the other binocular effects covered in this section, salient rivalry is likely to be universally considered as an undesirable visual artifact that compromises the visibility of the displayed content in AR.

2.5 Depth

It is well established that the visual system can use positional differences in the two eyes' images (binocular disparities) to infer depth information. It has been recently shown, however, that dichoptically tonemapped natural imagery with interocular contrast and luminance differences can generate a sense of depth as well [51, 60]. However, this depth effect has been elusive to vision science research, as it is harder to elicit consistently compared to binocular luster, rivalry, and stereoscopic depth (from binocular disparity). Psychophysical studies have demonstrated an anomalous depth effect (also referred to as the "sieve effect" and "rivaldepth") with anticorrelated images in which a white pixel in the left eye matches to a black pixel in the right eye, but there is no binocular disparity [22, 35, 39] (Figure 2(e)). There is individual variation in this anomalous depth effect, however, such that some participants can perceive a reversal in depth but not others [20, 44]. It also is highly dependent on the stimulus configuration [19, 20]. These depth effects may also be associated with luster and rivalry. For example, one small study found that these three effects could all be induced with the same amount of dichoptic

luminance difference by simply changing the stimulus size [41]. Depending on the use case, this depth effect may be an additional tool for display designers to enhance depth impressions, since people often underestimate the distance of objects simulated via near-eye displays [12]. On the other hand, any anomalous depth effects may also be problematic for tasks that require fine depth accuracy.

2.6 Modeling Dichoptic Percepts

Considering the importance of the perceptual appearance of dichoptic imagery for display design, it would be useful to be able to predict binocular appearance given any pair of input images for the two eyes. Efforts have been made to develop models to predict various aspects of binocular percepts, but no model exists yet that has been shown to reliably predict a range of perceptual factors at once. For example, some models of binocular combination focus on implementing the mechanisms of early stages of interocular interaction (e.g., interocular suppression) based on basic stimulus properties (e.g., contrast) [11, 15, 24, 32], while other models, particularly those focused on rivalry, employ higher-level frameworks such as perceptual inference and decision making [6, 21]. However, oftentimes these perceptual models intend to predict only a single aspect of appearance, and most prior work has focused on using controlled stimuli targeted to elicit one type of effect only. Some prior work has explored the perceived image quality of natural dichoptic images, as an extension of conventional 2D image quality metrics [5, 56]. Recent approaches in this domain have incorporated models of binocular processing; however, the evaluations focus on predicting a single dimensional measure of 3D image quality [7, 13, 45]. To support models that can predict the multi-faceted appearance of dichoptic stimuli, a better understanding of how multiple perceptual effects might co-occur is needed.

3 PERCEPTUAL EXPERIMENTS

In this article, we present the results of two perceptual experiments designed to examine all five of the aforementioned perceptual factors in dichoptic appearance together. We aim to provide a more holistic picture of what dichoptic stimuli may look like to users and in this way inform display design decisions. For example, it would be beneficial to know if there is any systematic relationship between the different perceptual effects. Are different effects associated more or less with different amounts of interocular image differences? Is there a “sweet spot” for optimal user experience where perceptual artifacts like binocular rivalry are minimized but the sense of depth or contrast is enhanced? How does the perceptual outcome change when viewing different spatial patterns?

In Experiment 1, we examine how spatial complexity and interocular contrast differences influence the occurrence of the different perceptual effects. This experiment uses conventional psychophysical stimuli. It aims to validate our multiquestion experimental procedure and understand the potential relationships between the perceptual factors of interest. Experiment 1 was conducted as part of a larger psychophysical study, and some non-overlapping results from this study were already reported in [52]. In Experiment 2, we leverage the paradigm from Experiment 1 to more directly examine how dichoptic imagery varies in appearance in optical see-through AR scenarios. We simulated stimuli in which the AR content is brighter in one eye than the other, which results in both interocular differences in luminance and contrast. Contrast for AR content was defined as the ratio of the maximum AR luminance over the maximum luminance of the background.

3.1 Participants

Two groups of 34 adults participated in Experiment 1 (23 females, ages 19–32 years) and Experiment 2 (25 females, ages 18–34 years). All participants had normal or corrected-to-normal visual acuity and normal stereo vision (measured with the Randot Stereotest). The experimental procedure was approved by the Institutional Review Board at University of California, Berkeley, and all participants gave informed consent prior to beginning the study.

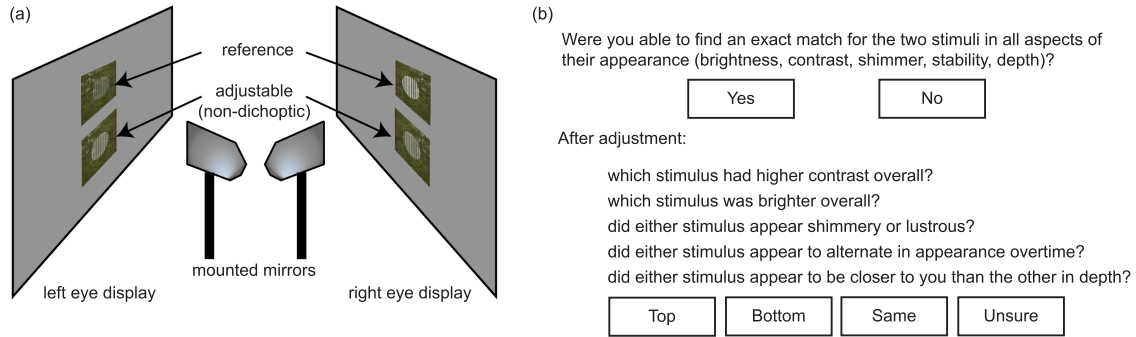


Fig. 3. (a) Experimental setup, in which two stimuli were shown to participants in a mirror haploscope. At the start of each trial, participants were asked to match the appearance of the two stimulus targets (e.g., the circular pattern) as best they could by varying the adjustable stimulus (natural image credit: © October 2016 SYNS Dataset [1]). (b) Following the matching task, a set of follow-up questions and response options (boxed) were shown on the screen for the participants to select based on what they saw during the matching phase.

3.2 Experimental Setup

Stimuli were displayed on a desk-mounted mirror haploscope (Figure 3(a)) to allow for independent presentation of images to the left and right eyes (presented on two LG 32UD99-W LCD displays). This system enabled precise control over the stimulus appearance in each eye without potential interference of optical imperfections that are common in wearable systems (e.g., optical distortions or vignetting). The viewing distance was 63 cm, and participants were head-fixed with a chin rest. The spatial resolution of each display was $3,840 \times 2,160$ pixels per eye (~ 60 pixels per visual degree). The experiment room was dark during the experiment.

To calibrate the displays' luminance, we used a PR650 spectrophotometer to measure the maximum white of each display. Then, we manually adjusted the brightness settings of the displays to achieve the best possible match for the maximum luminance. This adjustment resulted in a maximum luminance of 168 cd/m^2 for the left eye (white point $(x, y) = 0.31, 0.31$) and 164 cd/m^2 for the right eye (white point $(x, y) = 0.32, 0.32$). We then empirically measured the gamma nonlinearity of each display so that we could adjust the brightness and contrast of all stimuli in units that were linear with respect to the luminance output. We determined the grayscale gamma nonlinearity of each display perceptually using dithering, and generated a look-up table with the same gamma correction applied to each of the RGB channels. The resulting mid-gray luminance was verified with the spectrophotometer to be approximately half of the maximum luminance and was within 8 cd/m^2 between the two displays. The match between the two displays is also supported by the fact that the secondary perceptual effects such as luster and rivalry were almost never reported for the non-dichoptic stimuli in the experiment.

All stimuli in Experiment 2 were shown in standard sRGB colorspace. While this color gamut is likely representative of typical AR imagery, it has a more limited chromatic range than real natural environments. We adopt a standard of representing light levels in units of linearized pixel intensity in which the minimum light level of the display is assigned a value of 0 and the maximum is assigned a value of 1. With LCD displays, however, some light is always emitted from the display panel backlight, even when the pixel levels are set to 0.

3.3 Task

In a series of trials, participants were presented with pairs of stimuli to compare. One stimulus was presented on the top half of the screens and the other on the bottom half. One stimulus was identical in the two eyes (non-dichoptic) and the other stimulus (usually) comprised a dichoptic pair as described below. We call this latter stimulus the *reference*. Participants used keyboard presses to adjust the contrast of a target pattern in the

non-dichoptic stimulus to match the appearance of the target in the reference stimulus as best as they could (Figure 3(a)). They could look back and forth between the stimuli and could spend as much time as they needed to obtain the best match. The positions of the reference stimulus and the adjustable non-dichoptic stimulus were swapped for half of the participants, meaning that half of the participants saw the reference stimulus always on the top and the other half saw it always on the bottom.

After participants indicated that they had found the best match, the stimuli disappeared and they were shown several prompts to assess which, if any, perceptual differences there were between the reference and their best match (Figure 3(b)). They were first asked whether they were able to find an exact match or not. If the answer was no, they were asked to judge the contrast, brightness, luster, rivalry, and depth of their best match against the reference stimulus. The prompts shown in Figure 3(b) were presented sequentially on the screen. Response options to each prompt were top, bottom, same, and unsure. Responses of “top” or “bottom” indicated which stimulus was associated with the stronger perceptual effect. Based on pilot testing, we selected wording to describe luster and rivalry that best matched how participants described these effects (third and fourth questions, respectively). For the luster, rivalry, and depth questions, people were instructed to use the response option “same” when neither stimulus had the effect. Prior to starting the experiment, participants were shown images to help them understand what was meant by rivalry and luster. We showed them orthogonally oriented gratings in each eye to demonstrate how binocular rivalry looks. A square stimulus with different shades of gray in each eye was used to explain what luster looks like. Participants also completed 10 practice trials to get familiar with the task.

3.4 Stimuli

In Experiment 1, we used grayscale pattern stimuli to probe the nature of people’s responses to the visual appearance questions. In Experiment 2, we used stimuli designed to mimic the appearance of optical see-through AR systems.

3.4.1 Experiment 1. We used two common types of vision research stimuli in this experiment: vertical sine wave gratings with a spatial frequency of 5 cycles-per-degree (cpd) and a 1/f (“pink”) noise pattern with a broad frequency amplitude spectrum similar to that of natural images (Figure 4(a)). In a previous experiment, we found differences between the dichoptic contrast percepts of these grating and noise stimuli [52]. Therefore, in addition to these two stimuli we also included three intermediate noise patterns that shared some similarities with the grating patterns (also shown in Figure 4(a)): we matched the pixel intensity distribution of the 1/f noise pattern to the grating through histogram matching (histogram-matched), we bandpass-filtered the 1/f noise image and only kept spatial frequencies between 4 and 6 cpd (5 cpd bandpass), and we repeated the first row of the 1/f noise image for all rows to create a broadband vertical grating (broadband). Each stimulus image was 8-bit and spanned the full range of 0–255 bit levels.

Under realistic viewing conditions, targets of visual inspection are rarely viewed in isolation (i.e., against a uniform background). Instead, the surrounding context provides additional visual information that may play a role in determining the appearance. Thus, each image of the stimulus consisted of a 2° circular target of interest embedded in a binocular 4° by 4° surround region with the same type of spatial pattern. For example, the grating target was embedded in a grating pattern with the same spatial frequency and orientation (Figure 4(b)), whereas the 1/f noise was embedded in 1/f noise. To vary the contrast of each eye’s target region, we normalized the image range from 0 to 1 and rescaled the values around the mean value as follows:

$$I = c(I_0 - \mu) + \mu, \quad (1)$$

where I_0 denotes the original image, μ denotes the mean pixel intensity of that image, I denotes the new image, and c is a scalar value that determines the amount of contrast reduction. To generate the reference stimuli, the contrast (c) of the target for the left and right eyes (c_L and c_R) was set to 0, 0.25, 0.5, or 1, resulting in 16 possible combinations between the two eyes (e.g., $c_L = 0.25$ and $c_R = 1$, $c_L = 1$ and $c_R = 0.5$). We did not present a stimulus

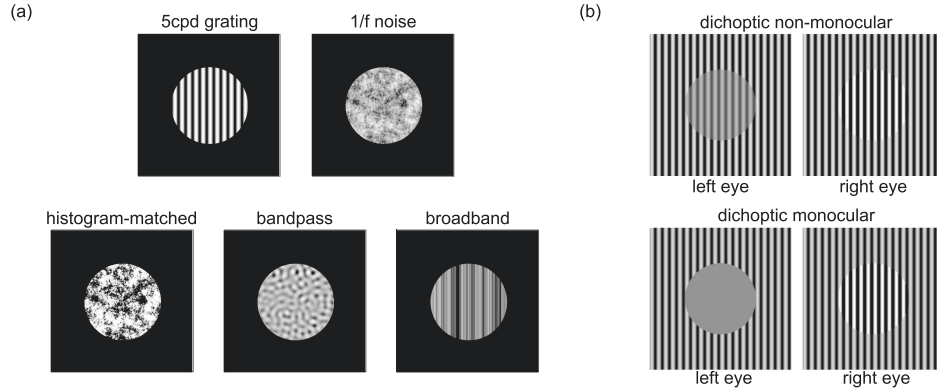


Fig. 4. (a) Five stimulus target patterns used in Experiment 1. (b) An example of two types of dichoptic references that were used: non-monocular reference stimuli had different contrast for each eye’s target and both eyes’ target contrasts were greater than 0, whereas monocular reference stimuli only had a target visible in one eye. Recall that all targets were embedded in a square surround region that matched the average contrast in the reference targets, and had the same type of spatial pattern as the targets.

with zero contrast in both eyes, so only 15 combinations were used. Of these, six combinations had $c = 0$ in one eye and $c > 0$ in the other eye, which we refer to as the special case of dichoptic stimuli with a monocular target. The other six dichoptic combinations were non-monocular (visible target in both eyes) (Figure 4(b)). The remaining three combinations resulted in non-dichoptic stimuli ($c_l = c_r$) that were used as control/catch trials. The contrast of the square outside of the target region was always equal to the average contrast of the two eyes’ target regions and non-dichoptic. All stimuli were shown on a uniform mid-gray background. In total, there were 75 trials (5 stimulus patterns \times 15 contrast combinations).

3.4.2 Experiment 2. We created stimuli that simulated AR visual experiences by compositing a virtual icon with a naturalistic background image. The virtual icon was then used as the target for the perceptual task. We tested four different patterns for the virtual icons. To have a baseline comparison with Experiment 1, we included the 5 cpd grating and 1/f noise pattern again. Based on the results of Experiment 1, we were interested in understanding if more realistic AR content would appear similar to the two baseline stimuli or not. We thus selected two different icon patterns from an existing library [48], which we refer to as simple and complex icons (Figure 5(a)). The grating, noise, simple, and complex icon stimuli were all overlaid on an image of a natural background from the SYNS dataset [1] (Figure 5(b)). The same background image was used for all icons to focus on the potential perceptual effects associated with each unique icon. Similar to Experiment 1, target regions (the icons) subtended 2° circles and the background region subtended a 4° square.

The contrast adjustment of the icons was performed similarly to Experiment 1, with some key differences to more closely simulate the joint contrast/luminance modulations that can occur when one display in an optical see-through AR system is brighter than the other. In particular, since these systems use additive light, we simulated the addition of the icon image onto the natural background. Pixel values of the background image (B) were scaled down by a factor of 2 so that only half of our display’s dynamic range was used to simulate the background and the other half could be used for the icons. This effectively provides a maximum AR contrast of 2:1 against the background. The normalized 8-bit, three-color channel icon images (A) were also downscaled by a factor of 2 before being multiplied by the different scale factors, such that the maximum normalized pixel value in the combined image was equal to 1:

$$I = c \left(\frac{A}{2} \right) + \left(\frac{B}{2} \right). \quad (2)$$

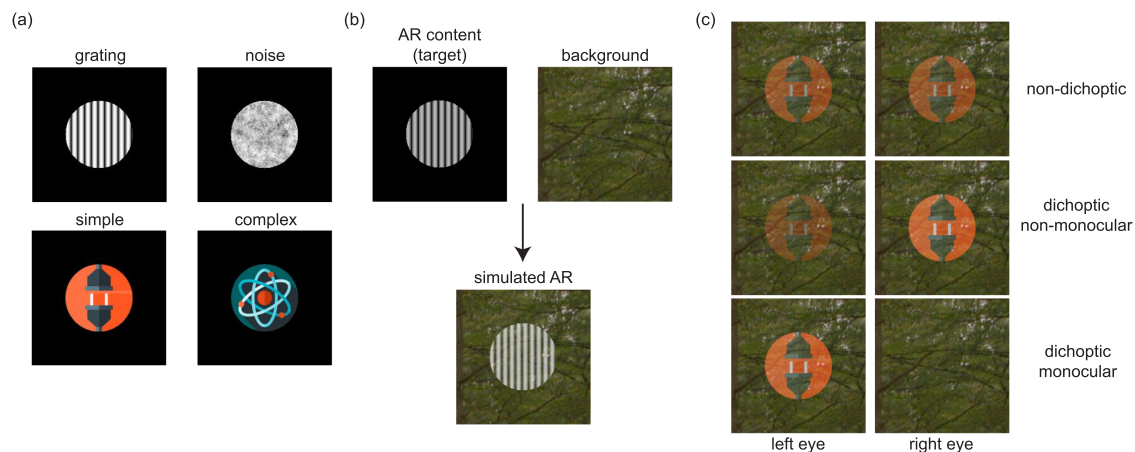


Fig. 5. (a) Four icon stimulus patterns used in Experiment 2. (b) We simulated the appearance of an AR target on a natural background by compositing each icon [48] with a forest scene (© October 2016 SYNS Dataset [1]). (c) The AR target in the reference stimulus could be non-dichoptic, dichoptic but non-monocular, or fully monocular.

All contrast adjustments were made in linear units based on the assumption that all color channels were encoded with a gamma non-linearity of 0.45 (e.g., normalized bit values from the background and icon images were exponentiated to $1/0.45$ prior to being combined). We used the same contrast combinations as described for Experiment 1 for the AR target in this experiment (Figure 5(c)). The surround region was identical in the two eyes. In total, there were 120 trials in Experiment 2 (4 stimulus conditions \times 15 luminance combinations \times 2 repeats).

3.4.3 Catch Trials. On some trials in both experiments, we presented a non-dichoptic reference to check that participants were following the instructions. We used the matching performance during these trials to exclude participants who were not performing the task reliably. Two participants from Experiment 1 and three participants from Experiment 2 were excluded because their matching error exceeded 1.5 times the interquartile range of all participants' errors. For the results presented below, $N = 32$ for Experiment 1 and $N = 31$ for Experiment 2.

3.5 Statistical Analyses

For the contrast matching task, the results for which non-dichoptic stimulus produced the closest perceptual match to the reference stimulus were fitted with a standard binocular contrast combination model. The model assumed that the binocular contrast percept was a weighted average of the contrast shown to the left and right eyes (see Figure 8(a)). The weights for the two eyes were constrained to sum to one, such that this model contained only one free parameter. Best fitting weights for each participant were determined with a grid search that minimized the square root of the mean squared error between the data and the model prediction across the trials. An analysis of the estimated weights for the stimuli in Experiment 1, in combination with additional experiments and stimuli, were previously reported elsewhere [52] and are summarized briefly in the following Results section. The estimated weights for Experiment 2 are reported here, and were analyzed with two one-way ANOVAs to examine the effects of stimulus type and interocular difference. Follow-up pairwise comparisons were done using t -tests with Bonferroni correction.

For the perceptual questions following the contrast matching task, we used mixed-effect logistic regression models to fit the responses and evaluate which stimulus properties were associated with different perceptual reports, with participants modeled as random intercepts. For each analysis, we include tables that report the coefficients, 95% confidence intervals, t statistics, and p values associated with a set of stimulus properties modeled

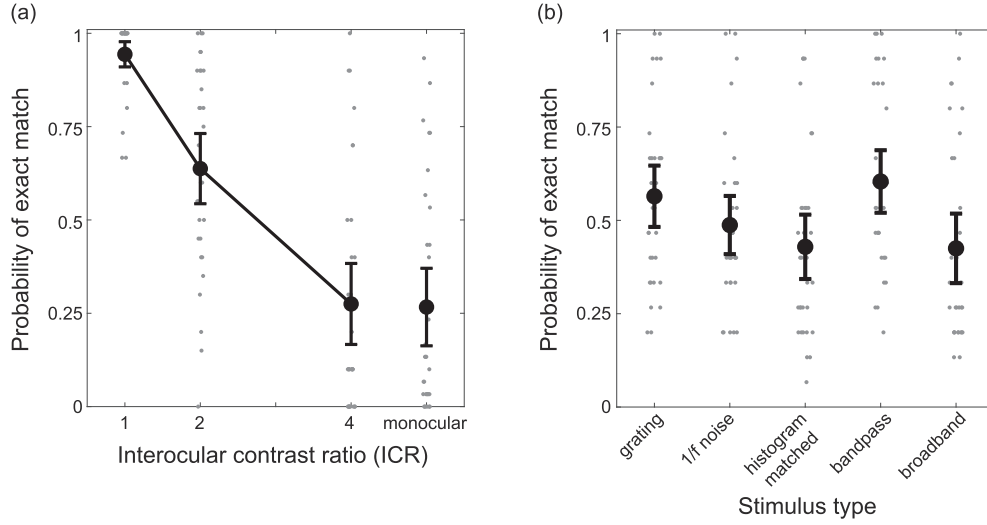


Fig. 6. Results for the exact match question in Experiment 1. Large black dots represent the average probability of finding an exact perceptual match across participants. Error bars are 95% confidence intervals. The smaller gray dots represent each participant's data. (a) The probability of exact matches across all interocular contrast ratios (ICR) in the two eyes, including monocular targets. (b) The probability of exact matches for each stimulus pattern with all ICRs included, including monocular trials.

as fixed effects. A qualitative examination of the data did not suggest that any notable interactions were present, so for simplicity we do not investigate or report interactions. For some analyses, we use a separate model to examine the difference between responses to monocular targets and other dichoptic stimuli so that we can treat monocular versus non-monocular targets as a categorical predictor.

4 RESULTS

4.1 Experiment 1

4.1.1 Contrast Matching. The contrast matching results from this experiment were already reported in detail elsewhere [52]. In brief, for most stimulus types people tended to match the non-dichoptic stimulus to the higher contrast image seen by either the left or the right eye. That is, the higher contrast image dominated binocular perception in a close to winner-take-all fashion. However, the individual variability was high for the 5 cpd grating stimulus in particular: for this stimulus, some participants' data were more consistent with simple averaging or even a loser-take-all pattern in which the lower contrast stimulus dominated the binocular percept. These results highlight the possibility that binocular contrast percepts may vary depending on the stimulus properties, which we will return to in the analysis of the contrast matching results for Experiment 2.

4.1.2 Probability of Finding an Exact Match. The data indicated that the best perceptual match participants could find was not always an *exact* perceptual match. The probability that participants could find an exact perceptual match to the reference stimulus varied systematically as a function of the interocular contrast difference and the stimulus pattern, although there was substantial individual variation (Figure 6). Figure 6(a) shows how the magnitude of the contrast difference between the two eyes was associated with dramatic changes in the probability of a perceptual match. To characterize the contrast differences, we use the ratio of the higher contrast target to the lower contrast target, which we call the **interocular contrast ratio (ICR)**. Our rationale is that human vision tends to follow Weber's Law—for example, the amount of luminance difference required to detect

Table 1. Logistic Regression Model for Experiment 1 Exact Match Question

Experiment 1	Coefficient (95% CI)	<i>t</i>	<i>p</i>
Intercept	5.13 (4.27, 5.99)	11.74	<0.001*
ICR	-1.62 (-1.81, -1.44)	-17.37	<0.001*
1/f noise	-0.46 (-0.95, 0.02)	-1.84	0.06
Histogram matched	-1.05 (-1.53, -0.56)	-4.28	<0.001*
Bandpass	0.56 (0.04, 1.08)	2.10	0.04*
Broadband	-1.31 (-1.79, -0.83)	-5.36	<0.001*

ICR is a continuous variable and stimulus types are categorical predictors (with grating used as the baseline). The coefficient reflects an increase or decrease in the probability that an exact match was obtained. Positive values indicate more exact matches, and negative values indicate fewer exact matches. Coefficients that are significantly different from zero based on the *t*-statistics (degrees of freedom = 1,434) are marked with asterisks (*).

Table 2. Logistic Regression Model for Experiment 1 Comparing Dichoptic Trials with Non-monocular Targets (ICR = 4) to Trials with Monocular Targets

Experiment 1	Coefficient (95% CI)	<i>t</i>	<i>p</i>
Intercept	-1.70 (-2.51, -0.88)	-4.08	<0.001*
Monocular	-0.07 (-0.45, 0.30)	-0.39	0.70

The non-monocular trials were used as the baseline. Data are reported in the same format as Table 1 (degrees of freedom = 1,278).

a luminance change is proportional to the background luminance—and as such, this ratio is likely to reflect the salience of the contrast differences in our stimuli [14, 60].

An ICR of 1 means that the reference stimulus had the same target contrast in each eye and was non-dichoptic. As expected, participants were able to find an exact perceptual match close to 100% of the time when this was the case. A larger ratio indicates a larger contrast difference between the two eyes (i.e., an ICR of 4 means one eye's contrast is four times the contrast of the other eye). As ICR increased from 1 to 4, participants were on average less likely to find an exact match, with only about a quarter of the stimuli resulting in an exact match when the ICR was equal to 4. We ran a logistic regression model with ICR (excluding monocular trials) and stimulus type as regressors. We can take the coefficients from the regression model (Table 1) and exponentiate them to obtain the odds ratios for the predictors. The coefficient of -1.62 for the ICR, for example, translates to an odds ratio of 0.20, meaning that for each one-unit increase in ICR, the odds of getting an exact match is 0.20 times less.

Next, we examine the results when the reference stimulus was monocular. Monocular reference stimuli, in which one eye had a target contrast of 0 (i.e., uniform gray embedded in a binocular surround region), have an ICR of infinity (Figure 6(a), labeled as monocular). For these stimuli, we ran a separate regression model containing a categorical predictor on a subset of the data, comparing just the monocular trials to the non-monocular trials with an ICR of 4. The results suggest that the probability of finding an exact perceptual match was not notably lower for monocular stimuli as compared to dichoptic stimuli with a large ICR (Table 2).

The probability of finding exact matches was less affected by stimulus type (Figure 6(b), Table 1). For this analysis, we used the grating stimulus as the baseline and examined the odds associated with the four other stimulus types. We found that the probability associated with the grating stimulus was not significantly different from the 1/f noise stimulus, but was significantly different from all three intermediate patterns. Compared to the grating stimulus, the odds ratio for the histogram-matched noise, bandpass noise, and broadband grating stimuli

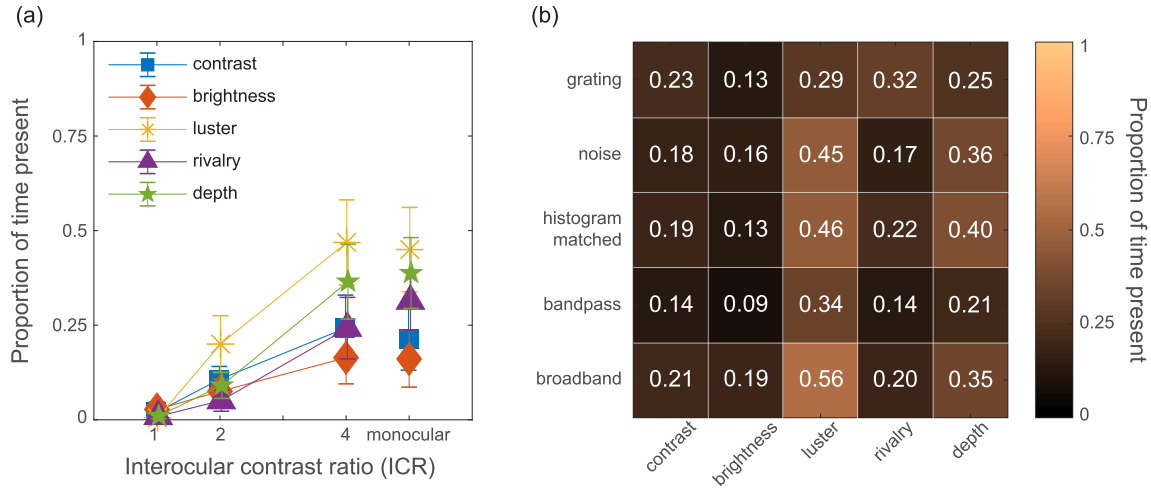


Fig. 7. Results for the five perceptual effects measured in Experiment 1. (a) The average proportion of trials across participants (with 95% confidence interval) in which each of the effects was present as a function of interocular contrast ratio (ICR), and for monocular targets. (b) Heatmap showing the average proportion of time that each effect (x axis) was present for each stimulus type (y axis) across all dichoptic trials (ICR = 2, 4, monocular).

were 0.35, 1.74, and 0.27, respectively. This result suggests that the spatial pattern of the stimulus may influence the chances that people see phenomena like luster, rivalry, and depth in dichoptic stimuli, but the effect appears to be smaller compared to the effect of ICR.

4.1.3 Perceptual Appearance of Dichoptic Stimuli. What perceptual effects did people experience when they were unable to find an exact perceptual match by varying stimulus contrast, and how do these perceptual effects vary across different interocular contrast ratios and stimulus types? To answer these questions, we next look at participants' responses to the follow-up questions about perceived contrast, brightness, luster, rivalry, and depth.

First, as a sanity check, we calculated which stimulus participants reported seeing luster and rivalry in. We expected participants to select the dichoptic reference stimulus as the one that elicits these perceptual phenomenon, because the adjustable stimulus was always non-dichoptic and should not elicit luster or rivalry. The results were consistent with this expectation. When luster was detected in one of the stimuli, the reference stimulus was selected 98% of the time. For rivalry, it was 94%. When a depth difference was detected, participants also tended to indicate that the dichoptic stimulus was closer (84% of the time). For the brightness and contrast questions, we did not expect participants to systematically select either stimulus because we do not have a strong hypothesis that dichoptic stimuli should appear systematically higher or lower in contrast or brightness than non-dichoptic ones. Indeed, the choices for these prompts were closer to chance (56% and 61% of the time, respectively).

For the main analysis, we re-coded the data to simply indicate whether people perceived a difference or not for each perceptual factor. When participants made any response other than "same" for a given prompt, a perceptual difference was considered to be present. The average percentage of "unsure" responses across all the prompts was low (mean = 1.19%, standard error = 0.19%, median = 0.53% of all responses) and similar across all questions, and the results do not notably change if we omit these responses.

When the reference was non-dichoptic (ICR = 1), there were minimal perceptual differences, as expected from the analysis of exact matches (Figure 7(a)). That is, on these trials participants were unlikely to indicate any perceptual differences between the two stimuli. As the ICR increased, all five effects started to become more noticeable. The most common perceptual differences across all ICR levels were binocular luster, depth, and rivalry,

Table 3. Logistic Regression Models for the Occurrence of Perceptual Effects in Experiment 1

Expt. 1	Coefficient (95% CI)	<i>t</i>	<i>p</i>	Expt. 1	Coefficient (95% CI)	<i>t</i>	<i>p</i>
Contrast				Contrast			
Intercept	-4.41 (-5.07, -3.76)	-13.20	<0.001*	Intercept	-1.48 (-2.07, -0.88)	-4.86	<0.001*
ICR	0.77 (0.61, 0.93)	9.62	<0.001*	Monocular	-0.27 (-0.63, 0.08)	-1.51	0.13
Brightness				Brightness			
Intercept	-4.46 (-5.14, -3.78)	-12.80	<0.001*	Intercept	-2.25 (-2.95, -1.56)	-6.37	<0.001*
ICR	0.63 (0.46, 0.80)	7.13	<0.001*	Monocular	0.01 (-0.40, 0.42)	0.05	0.96
Luster				Luster			
Intercept	-5.27 (-6.13, -4.41)	-12.02	<0.001*	Intercept	0.13 (-0.68, 0.95)	0.32	0.75
ICR	1.42 (1.25, 1.60)	15.59	<0.001*	Monocular	-0.20 (-0.54, 0.13)	-1.18	0.24
Rivalry				Rivalry			
Intercept	-6.65 (-7.65, -5.65)	-13.04	<0.001*	Intercept	-1.63 (-2.23, -1.02)	-5.25	<0.001*
ICR	1.26 (1.02, 1.49)	10.46	<0.001*	Monocular	0.51 (0.17, 0.86)	2.91	0.004*
Depth				Depth			
Intercept	-5.59 (-6.35, -4.82)	-14.40	<0.001*	Intercept	-0.61 (-1.36, 0.13)	-1.61	0.11
ICR	1.29 (1.12, 1.46)	14.65	<0.001*	Monocular	-0.02 (-0.33, 0.29)	0.12	0.90

Left: Each effect's occurrence as a function of ICR (ICRs of 1, 2, and 4; degrees of freedom = 1,438). Right: Each effect's difference in occurrence between non-monocular (ICR = 4) and monocular dichoptic trials (degrees of freedom = 1,278). Data are reported in the same format as Tables 1 and 2.

in the order from most likely to less likely. The results also suggest that different effects tended to co-occur to some extent, because the proportions for high ICR trials sum to a value greater than 1. Indeed, experiences of these perceptual phenomena were not mutually exclusive. Across all participants, the mean number of perceptual differences per dichoptic trial was greater than 1, with marginal statistical significance (mean = 1.27, median = 1.17, $t(31) = 1.94$, $p = 0.06$), and this amount increased with increasing ICR (e.g., the mean and median were 1.60 and 1.40 for an ICR of 4, $t(31) = 3.38$, $p = 0.002$).

We used five logistic regressions to examine the occurrence of each perceptual effect separately. Table 3 (left) shows the association between ICR (ICRs of 1–4) and the presence of each perceptual effect. The ICR coefficients for all effects were positive and statistically significant, suggesting that the occurrence of all perceptual effects increased systematically as ICR increased. Based on the magnitude of the ICR coefficients, luster had the largest increase. In terms of odds ratios, we observed about a factor of 4.2 increase in the odds of luster for each one-unit increase in ICR, as compared to a 3.5 increase in rivalry, 3.6 in depth effects, and 2.2 and 1.9 increases in odds of contrast and brightness effects, respectively.

Next, we directly compared the trials with a monocular target to trials with a non-monocular high ICR target (ICR = 4) (Table 3, right). The fits to this subset of trials indicate that the monocular targets were associated with a relative increase in binocular rivalry (odds ratio of 1.67), while no other effects were notably different.

Lastly, we looked qualitatively at how the perceptual effects differed among different stimulus patterns. Figure 7(b) shows the occurrence of perceptual differences for each stimulus type out of all the dichoptic trials (i.e., all trials except when the ICR was equal to 1). The lighter the color in the matrix, the higher the likelihood that there was a difference associated with each effect (x-axis label) for the given stimulus pattern (y-axis label). The results suggest that different stimulus patterns may have a different set of perceptual effects. For example, the grating stimulus had fewer perceptual differences overall, and a slightly higher rate of rivalry than luster. The more complex patterns were associated with relatively higher rates of luster, all of which exceeded the occurrence of rivalry. Taken together, this set of results suggests that rivalry may be a concern particularly for monocular stimuli and for simple grating stimuli. These results serve to highlight the importance of

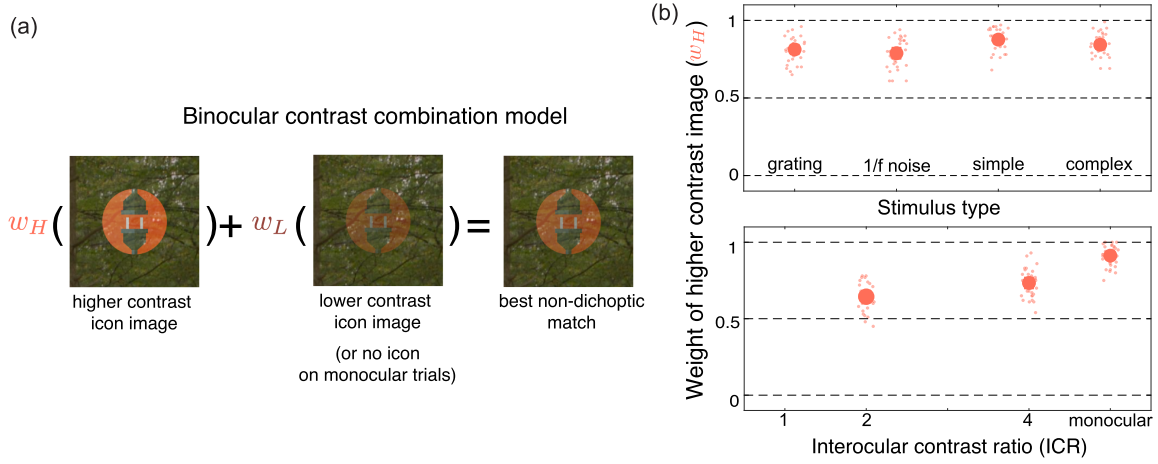


Fig. 8. (a) Schematic of our simple weighted combination model used to quantify binocular contrast perception for the contrast matching results in Experiment 2. Image credits: icon [48], background © October 2016 SYNS Dataset [1]. (b) The matching result is expressed as the weight for the high-contrast eye (w_H) across different stimulus types (top) and different dichoptic conditions (bottom). The large dots represent the average weights across all participants, and the small dots represent each participant's fitted weight. The 95% confidence interval for each average is either smaller than or approximately the same size as the circular marker.

investigating these perceptual effects using visual stimuli that mimic the visual appearance of genuine AR experiences, which we will describe in the next section.

4.2 Experiment 2

The stimuli used in Experiment 2 were designed to more closely mimic the visual experience of optical see-through AR, with natural backgrounds, partially transparent imagery, and a coupling of contrast changes with stimulus brightness.

4.2.1 Contrast Matching. First, we looked at the contrast matching results. We fitted the contrast matching data using a simple weighted combination model where the weights for the high and low contrast eye add up to 1 (Figure 8(a)). In Figure 8(b), the weights assigned to the higher contrast stimulus across all trials for the different stimulus types are shown in the top panel, and the weights for the higher contrast stimulus for different ICRs (except ICR = 1 where the model is unconstrained) across all stimulus types are shown in the bottom panel.

The results for the different stimulus patterns are all generally consistent with previously published results, in which the higher contrast stimulus dominates (the weight on the higher contrast image was near 1, approximating a winner-take-all binocular combination rule). However, a one-way ANOVA showed that there were significant differences among the different stimulus types for the weights ($F(3, 90) = 26.28, p < 0.001$). Follow-up pairwise t -tests revealed all pairs of stimulus types were significantly different from each other, except for the grating and noise patterns (Table 4, left). This suggests that stimulus pattern plays a significant role in determining the bias in binocular combination, although the average weight across stimulus types was always greater than 0.78.

The effect of ICR on contrast matching was also significant ($F(2, 60) = 408.73, p < 0.001$). All levels of ICR were significantly different from each other, suggesting that ICR has a robust influence on the perceived binocular contrast (Table 4, right). Importantly, monocular trials were associated with significantly greater high-contrast weights than the other conditions, showing evidence of Fechner's paradox for AR stimuli. Indeed,

Table 4. Pairwise t -Test Results for Difference in Weights Across Different Stimulus Patterns and Different ICR Levels in Experiment 2

Experiment 2	t	p	d
grating-noise	2.61	0.09	0.27
grating-simple	-6.01	<0.001*	-0.82
grating-complex	-2.91	0.04*	-0.38
noise-simple	-7.55	<0.001*	-1.02
noise-complex	-5.02	<0.001*	-0.63
simple-complex	3.34	0.01*	0.44

Experiment 2	t	p	d
ICR 2-4	-10.4	<0.001*	-0.97
ICR 2-monocular	-28.3	<0.001*	-3.35
ICR 4-monocular	-17.0	<0.001*	-2.19

The test statistics (degrees of freedom = 30), Bonferroni corrected p values, and effect sizes (Cohen's d) are reported.

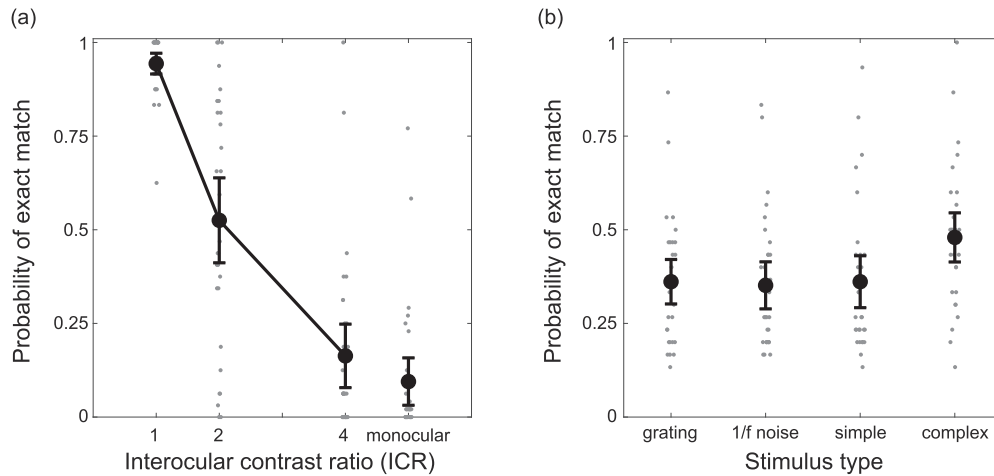


Fig. 9. Results for the exact match question in Experiment 2. Large black dots represent the average probability of finding an exact perceptual match across subjects. Error bars are 95% confidence intervals. The smaller gray dots represent each participant's data. (a) The probability of exact matches across different interocular contrast ratios (ICR) in the two eyes, and for monocular targets. (b) The probability of exact matches for each stimulus pattern with all ICRs included, including monocular trials.

when the ICR was the lowest dichoptic value (2) the binocular contrast combination was closer to averaging than winner-take-all.

4.2.2 Probability of Finding an Exact Match. The effect of ICR on the probability of finding an exact perceptual match was similar to Experiment 1 for the AR-like stimuli used in Experiment 2 (Figure 9(a), Table 5). In this experiment, the probability of finding exact matches when the contrast ratio was high (ICR = 4) or when the stimulus was monocular was quite low. For each unit increase in ICR, the odds of getting an exact match were 0.12 times less. For the four AR icon patterns used in this experiment, only the complex icon condition was associated with a match probability that was significantly different from the grating baseline, and this modulation was again substantially less than the differences associated with ICR (Figure 9(b), Table 5). When comparing the trials with a monocular AR target against the trials with a target ICR of 4, we found the monocular AR was associated with a significantly lower probability of finding an exact match (Table 6), with an odds ratio of 0.37.

Taken together, these results replicate and extend the findings from Experiment 1. The results indicate that there is substantial variation in the appearance of dichoptic AR stimuli that differ in interocular contrast, and that these appearances are subtly but lawfully modulated by the stimulus pattern.

Table 5. Logistic Regression Model for Experiment 2 Exact Match Question

Experiment 2	Coefficient (95% CI)	<i>t</i>	<i>p</i>
Intercept	4.82 (3.99, 5.66)	11.29	<0.001*
ICR	-2.12 (-2.31, -1.93)	-21.74	<0.001*
1/f noise	-0.20 (-0.54, 0.14)	-1.14	0.26
Simple	-0.18 (-0.52, 0.16)	-1.05	0.29
Complex	1.52 (1.15, 1.91)	7.86	<0.001*

ICR is a continuous variable and stimulus types are categorical predictors (with grating used as the baseline). The coefficient reflects an increase or decrease in the probability that an exact match was obtained. Positive values indicate more exact matches, and negative values indicate fewer exact matches. Coefficients that are significantly different from zero based on the *t*-statistics (degrees of freedom = 2,227) are marked with asterisks (*).

Table 6. Logistic Regression Model for Comparing Dichoptic Trials with Non-monocular Targets (ICR = 4) to Trials with Monocular Targets (Degrees of Freedom = 1,982)

Experiment 2	Coefficient (95% CI)	<i>t</i>	<i>p</i>
Intercept	-2.67 (-3.47, -1.87)	-6.54	<0.001*
Monocular	-0.98 (-1.36, -0.61)	-5.09	<0.001*

The non-monocular trials were used as the baseline. Data are reported in the same format as Table 5.

4.2.3 Perceptual Appearance of Dichoptic Stimuli. We performed a set of analyses on the perceptual appearance responses mirroring those described for Experiment 1. Similar to Experiment 1, the response patterns for luster and rivalry fit our expectations: when these effects were present, participants indicated that they saw them in the dichoptic reference 98% of the time. When a depth difference was detected, people indicated that the dichoptic stimulus was closer 98% of the time. For the brightness and contrast questions people selected the dichoptic stimulus as higher contrast or brighter 65% and 74% of the time, respectively. Again, we coded an effect as not present if participants responded with “same,” and present if they responded with one of the other three options. The number of “unsure” responses per participant was again low (mean = 0.78%, standard error = 0.10%, median = 0.33% of responses) and similar across questions. As in Experiment 1, recoding these responses did not change the pattern in the results.

When looking at whether or not effects co-occurred, the average number of perceptual effects per dichoptic trial across all participants was significantly greater than 1 (mean = 2, median = 1.82, $t(30) = 6.72$, $p < 0.001$), and this amount increased with increasing ICR (e.g., the mean and median were 2.54 and 2.42 for an ICR of 4, $t(30) = 6.36$, $p < 0.001$). Similar to Experiment 1, the probability that participants reported any perceptual effect increased as the stimulus ICR increased (Figure 10(a), Table 7, left). Luster was again the most commonly reported perceptual phenomenon associated with dichoptic imagery, and it increased the most with ICR. Across ICRs of 1, 2, and 4, there was a factor of 5.79 increase in the odds of luster for each unit increase in ICR, as compared to a 2.91 increase in rivalry, 3.62 in depth effects, 2.74 in contrast differences, and 2.91 in brightness differences.

Importantly, there were notable differences in the responses associated with AR targets that had high ICR and AR targets that were fully monocular (Table 7, right). The probability of all effects except for luster substantially increased for the monocular target compared with the non-monocular high ICR target. The odds ratios were 1.87 for contrast, 1.31 for brightness, 5.21 for rivalry, and 4.20 for depth. Qualitatively, the probability of reporting luster was lower for the monocular target, but this difference did not reach statistical significance.

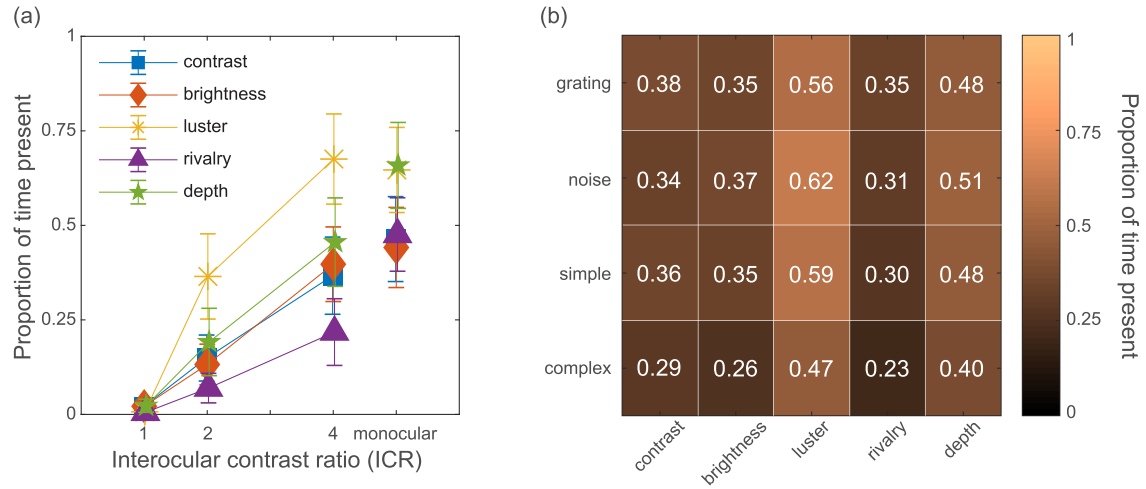


Fig. 10. Results for the five perceptual effects measured in Experiment 2. (a) The average proportion of trials across participants (with 95% confidence interval) in which each of the effects was present as a function of interocular contrast ratio (ICR), and for monocular targets. (b) Heatmap showing the average proportion of time that each effect (x axis) was present for each stimulus type (y axis) across all dichoptic trials (ICR = 2, 4, monocular).

Table 7. Logistic Regression Models for the Occurrence of Perceptual Effects in Experiment 2

Expt. 2	Coefficient (95% CI)	<i>t</i>	<i>p</i>	Expt. 2	Coefficient (95% CI)	<i>t</i>	<i>p</i>
Contrast				Contrast			
Intercept	-4.67 (-5.30, -4.04)	-14.54	<0.001*	Intercept	-0.79 (-1.53, -0.06)	-2.13	0.03*
ICR	1.01 (0.89, 1.13)	16.05	<0.001*	Monocular	0.63 (0.36, 0.89)	4.69	<0.001*
Brightness				Brightness			
Intercept	-4.69 (-5.23, -4.14)	-16.76	<0.001*	Intercept	-0.48 (-1.12, 0.16)	-1.48	0.14
ICR	1.07 (0.94, 1.19)	17.03	<0.001*	Monocular	0.27 (0.02, 0.52)	2.10	0.04*
Luster				Luster			
Intercept	-5.33 (-6.21, -4.44)	-11.83	<0.001*	Intercept	1.20 (0.41, 1.99)	2.99	0.003*
ICR	1.76 (1.59, 1.92)	20.87	<0.001*	Monocular	-0.22 (-0.50, 0.06)	-1.54	0.12
Rivalry				Rivalry			
Intercept	-6.01 (-6.80, -5.22)	-14.90	<0.001*	Intercept	-1.74 (-2.36, -1.12)	-5.54	<0.001*
ICR	1.07 (0.90, 1.23)	12.66	<0.001*	Monocular	1.65 (1.36, 1.94)	11.12	<0.001*
Depth				Depth			
Intercept	-5.31 (-6.13, -4.49)	-12.73	<0.001*	Intercept	-0.32 (-1.12, 0.48)	-0.78	0.44
ICR	1.29 (1.15, 1.43)	18.10	<0.001*	Monocular	1.43 (1.15, 1.72)	10.00	<0.001*

Left: Each effect's occurrence as a function of ICR (ICRs of 1, 2, and 4; degrees of freedom = 2,230). Right: Each effect's occurrence difference between non-monocular (ICR = 4) and monocular dichoptic trials (degrees of freedom = 1,982). Data are reported in the same format as Tables 5 and 6.

The association of each stimulus type with each perceptual difference is shown in Figure 10(b). Overall, there was no strong qualitative stimulus-dependent pattern. Unlike in Experiment 1, the grating stimulus was not associated with a unique pattern that deviated from the other stimuli when presented as a semi-transparent stimulus over a natural background. All stimulus types had binocular luster and depth differences as the predominant reported effects.

5 DISCUSSION

5.1 Predicting Perceptual Artifacts in Dichoptic AR Stimuli

Our results highlight the importance of understanding the multi-faceted nature of dichoptic percepts, particularly with visual stimuli that closely match genuine AR experiences. For example, with the simple stimuli used in Experiment 1, participants did not consistently report any of the dichoptic perceptual effects more than 50% of the time on average. But when we switched to AR-like stimuli in Experiment 2, we observed high rates of luster (about 75% on average in extreme dichoptic cases), along with a notable increase in rivalry for monocular AR-like stimuli.

We conducted a post hoc analysis using t -tests to compare the distributions of each of the perceptual effects in the two experiments. For this analysis, we focused on the highest interocular contrast ratio ($ICR = 4$) and the monocular conditions. We used an initial significance threshold of $p < 0.05$ and did not correct for multiple comparisons to avoid being overly conservative (a Bonferroni corrected p -value threshold for these comparisons would be $p < 0.005$). For the $ICR = 4$ condition, we found that brightness differences were significantly more prevalent in Experiment 2 as compared to Experiment 1 ($t(61) = 3.57, p < 0.001$, Cohen's $d = 0.90$). For the monocular condition, all effects except luster were more prevalent in Experiment 2: contrast ($t(61) = 3.50, p < 0.001$, Cohen's $d = 0.87$), brightness ($t(61) = 4.04, p < 0.001$, Cohen's $d = 1.01$), rivalry ($t(61) = 2.54, p = 0.01$, Cohen's $d = 0.63$), and depth ($t(61) = 3.12, p = 0.003$, Cohen's $d = 0.78$). We speculate that these differences may derive from a combination of the different spatial patterns of the AR stimuli used in the Experiment 2 and the more conventional stimuli used in Experiment 1 [52], as well as the fact that the AR stimuli had background content (natural foliage) that was partially visible behind the icons. However, the perceptual mechanisms that would modulate these effects in a stimulus-dependent manner are poorly understood.

Due to the differences between the two experiments, we propose that perceptually motivated guidelines for acceptable levels of ICR between two AR displays should be more conservative than might be assumed based on simpler stimuli. We can use the results from Experiment 2 to provide preliminary design guidelines for AR applications. As an example, we consider a case in which we want to adopt a strict threshold on the probability that a dichoptic stimulus contains any perceptual effects that deviate from a comparison non-dichoptic stimulus. Collapsing across all of the stimulus types (i.e., removing them as model parameters) and refitting the trial-by-trial data for the exact match question with a logistic regression model, we come to the following equation:

$$ICR_{max} = \frac{\ln\left(\frac{P}{1-P}\right) - 4.70}{-1.95}, \quad (3)$$

where P is the designer-selected minimum proportion of trials on which the dichoptic stimulus matches a non-dichoptic one (i.e., no perceptual effect), and ICR_{max} is the maximum acceptable ICR. For example, if the designer aims for a threshold of $P = 0.8$ (a perceptual match 80% of the time), they should aim for an ICR of no more than 1.7. However, this result reflects the data on average, and given the large individual variation in our data a more strict threshold may be appropriate to accommodate users who are more sensitive to dichoptic perceptual effects. For example, a 90% threshold would be associated with an ICR of 1.28 or less.

Currently, there is no publicly available dataset characterizing the typical binocular differences in contrast due to display defects and inefficiencies in AR systems. Prior work on optical eyebox limitations, however, suggests that interocular contrast differences in AR systems can span a broad range depending on the fit on the user's face and the movement of the device [4, 8, 43]. For example, systems with small eyeboxes may be particularly susceptible to large ICRs if one of the user's pupils moves to the edge of the eyebox. On the other hand, because our data show that observers are relatively tolerant to global differences in contrast between the two eyes, our results suggest a potential opportunity to reduce system power requirements by selectively attenuating the brightness of one display.

5.2 Modeling Interocular Differences and their Effect on Perception

The best metric for quantifying interocular differences in AR systems remains an open area of research. Here, we used the ratio between the overall contrast in each eye as the summary measure of interocular difference. However, there may be other metrics that could be more informative and practical. In particular, in Experiment 2 the stimuli differed in more than just contrast, so this metric is incomplete. Ideally, perceptual metrics of interocular differences should account for both luminance and contrast, and even color. For example, the luminance adjustment applied to the colored AR icons in Experiment 2 could also result in interocular color differences, especially when viewing monocular AR on a binocular background (e.g., a red monocular target against a green binocular forest background), which is known to elicit perceptual effects such as luster as well [33].

As AR technologies advance, the nature of the image quality problems and artifacts posed by these technologies will continue to change. Building better image-computable models of binocular combination will be crucial because these models can be used to develop metrics that account for arbitrary differences between the two eyes. However, the formulation of flexible models for perceived contrast in complex imagery, let alone binocular contrast perception, remains an ongoing area of research [18, 34, 37]. In our previous work, we explored how a Bayesian ideal observer model, which assumed binocular percepts are determined through a statistically optimal combination of binocular visual input and prior assumptions about the structure of the natural world, could explain specific properties of binocular depth perception [49]. However, this model did not account for any other properties of binocular appearance, like contrast, luster, and rivalry. Given the strong ability of Bayesian models—and related probabilistic cue combination models—to fit and predict perceptual phenomena [25], such approaches may be a fruitful way to formulate binocular perception more broadly. For example, it may be possible to predict the probability of perceived luster based on statistical regularities in the binocular differences created by lustrous metallic surfaces.

Generalizable models of binocular perception also have great appeal for developing tonemapping methods intended to improve image quality through binocular combination. For example, a recent line of research has looked at developing tonemapping methods that intentionally display different luminance and contrast information to the two eyes in order to improve the overall visual quality of a stereoscopic display system that cannot reproduce the full dynamic range of luminance found in the natural environment. Yet, at present the results of these approaches are mixed [51, 55, 58–60]. In our experiments, we found that people did often report contrast and brightness differences between the best match and the dichoptic reference stimulus. Furthermore, looking at the four response types, participants tended to select the dichoptic reference stimulus to be higher contrast and brighter in Experiment 2, suggesting the potential for dichoptic imagery to boost subjective image quality (the reference and the adjustable stimuli were about equally selected in Experiment 1). The current results do not reveal why participants may have experienced this “dichoptic boost” in Experiment 2, so this question represents a promising area for future research.

5.3 Potential Benefits of Dichoptic Contrast in AR

Here, we focused primarily on the potential negative consequences of interocular differences in display brightness/contrast for users of AR systems. However, some of the perceptual phenomena we characterized may be desirable. For example, the appearance that a dichoptic stimulus is closer in depth might be helpful for heads-up AR systems that display icons floating in front of the environment. However, we found that this depth effect generally co-exists with other phenomena and may be challenging to isolate. For example, as the likelihood of the depth effect increased, binocular rivalry also increased. Therefore, we did not find a “sweet spot” of interocular contrast differences where desired effects may dominate undesired ones. However, some effects were more readily detected at a lower interocular difference than others. In both experiments, binocular luster was more detectable than the other effects, and the good news is that rivalry remained relatively uncommon in comparison. This may be beneficial if designers want to leverage binocular luster to create a shiny metallic appearance

for a virtual object without rivalry effects. The exception to this observation was during monocular viewing, particularly in the AR-like situation simulated in Experiment 2. In this experiment, observers detected rivalry about half of the time during monocular trials, suggesting that even if a binocular display has large interocular differences, it may be preferable to a monocular system if rivalry is a concern. Lastly, we performed exploratory analyses to see whether some perceptual effects might be minimized if the higher contrast image in a binocular display was shown to the user's dominant eye. However, we did not find compelling evidence for eye dominance effects in the current dataset.

6 CONCLUSION

Binocular displays can introduce unwanted visual differences between the left and right eye's views. Here, we focused on the perceptual consequences of contrast differences for optical see-through AR systems in particular, but such interocular differences can occur in any binocular display system. Across two experiments, our results suggest that the binocular appearance of dichoptic imagery is multi-faceted, and the magnitude of the interocular difference between the two eyes is a main predictor for the intrusion of potentially detrimental perceptual effects such as luster and rivalry. Our study results provide an overview of supra-threshold perceptual effects, but understanding detection thresholds for these effects will provide valuable and complementary information for display design. As we continue to improve our understanding of the perceptual phenomena associated with binocular differences in AR devices, a careful consideration of both the scope and strength of these phenomena can help guide design choices that support a high-quality user experience.

ACKNOWLEDGMENT

The authors would like to thank Terrie Joo for helping with data collection.

REFERENCES

- [1] Wendy J. Adams, James H. Elder, Erich W. Graf, Julian Leyland, Arthur J. Lugtigheid, and Alexander Murry. 2016. The Southampton-York Natural Scenes (SYNS) dataset: Statistics of surface attitude. *Scientific Reports* 6 (2016), 35805.
- [2] Daniel H. Baker, Stuart A. Wallis, Mark A. Georgeson, and Tim S. Meese. 2012. Nonlinearities in the binocular combination of luminance and contrast. *Vision Research* 56 (2012), 1–9.
- [3] Randolph Blake. 2001. A primer on binocular rivalry, including current controversies. *Brain and Mind* 2 (2001), 5–38.
- [4] Ozan Cakmakci, David M. Hoffman, and Nikhil Balram. 2019. 31-4: Invited paper: 3D eyebox in augmented and virtual reality optics. *SID Symposium Digest of Technical Papers* 50, 1 (2019), 438–441.
- [5] Patrizio Campisi, Patrick Le Callet, and Enrico Marini. 2007. Stereoscopic images quality assessment. In *2007 15th European Signal Processing Conference*. 2110–2114.
- [6] Robin Cao, Alexander Pastukhov, Stepan Aleshin, Maurizio Mattia, and Jochen Braun. 2021. Binocular rivalry reveals an out-of-equilibrium neural dynamics suited for decision-making. *eLife* 10 (2021), e61581.
- [7] Zhibo Chen, Jiahua Xu, Chaoyi Lin, and Wei Zhou. 2020. Stereoscopic omnidirectional image quality assessment based on predictive coding theory. *IEEE Journal of Selected Topics in Signal Processing* 14, 1 (2020), 103–117.
- [8] Steven A. Cholewiak, Zeynep Başgöze, Ozan Cakmakci, David M. Hoffman, and Emily A. Cooper. 2020. A perceptual eyebox for near-eye displays. *Optics Express* 28, 25 (2020), 38008–38028.
- [9] Jian Ding, Stanley A. Klein, and Dennis M. Levi. 2013. Binocular combination of phase and contrast explained by a gain-control and gain-enhancement model. *Journal of Vision* 13, 2 (2013), 13–13.
- [10] Jian Ding and Dennis M. Levi. 2017. Binocular combination of luminance profiles. *Journal of Vision* 17, 13 (2017), 4–4.
- [11] Jian Ding and George Sperling. 2006. A gain-control theory of binocular combination. *Proceedings of the National Academy of Sciences of the United States of America* 103, 4 (2006), 1141–1146.
- [12] Fatima E. Jamiy and Ronald Marsh. 2019. Survey on depth perception in head mounted displays: Distance estimation in virtual reality, augmented reality, and mixed reality. *IET Image Processing* 13, 5 (2019), 707–712.
- [13] Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, and Christine Fernandez-Maloigne. 2017. Stereoscopic image quality assessment based on the binocular properties of the human visual system. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '17)*. 2037–2041.
- [14] Monika A. Formankiewicz and John D. Mollon. 2009. The psychophysics of detecting binocular discrepancies of luminance. *Vision Research* 49 (2009), 1929–1938.

- [15] Mark A. Georgeson, Stuart A. Wallis, Tim S. Meese, and Daniel H. Baker. 2016. Contrast and lustre: A model that accounts for eleven different forms of contrast discrimination in binocular vision. *Vision Research* 129 (2016), 98–118.
- [16] Bradley S. Gundlach, Michel Frising, Alireza Shahsafi, Gregory Vershbow, Chenghao Wan, Jad Salman, Bas Rokers, Laurent Lessard, and Mikhail A. Kats. 2018. Design considerations for the enhancement of human color vision by breaking binocular redundancy. *Scientific Reports* 8 (2018), 11971.
- [17] Nargess Hassani and Michael J. Murdoch. 2019. Investigating color appearance in optical see-through augmented reality. *Color Research & Application* 44, 4 (2019), 492–507.
- [18] Andrew M. Haun and Eli Peli. 2013. Perceived contrast in complex images. *Journal of Vision* 13, 13 (2013), 3–3.
- [19] Paul B. Hibbard and Jordi M. Asher. 2022. Robust natural depth for anticorrelated random dot stereogram for edge stimuli, but minimal reversed depth for embedded circular stimuli, irrespective of eccentricity. *PLoS One* 17, 9 (2022), 1–21.
- [20] Paul B. Hibbard, Kenneth C. Scott-Brown, Emma C. Haigh, and Melanie Adrain. 2014. Depth perception not found in human observers for static or dynamic anti-correlated random dot stereograms. *Plos One* 9 (2014), 1–9.
- [21] Jakob Hohwy, Andreas Roepstorff, and Karl J. Friston. 2008. Predictive coding explains binocular rivalry: An epistemological review. *Cognition* 108 (2008), 687–701.
- [22] Ian P. Howard. 1995. Depth from binocular rivalry without spatial disparity. *Perception* 24 (1995), 67–74.
- [23] Chang-Bing Huang, Jiawei Zhou, Yifeng Zhou, and Zhong-Lin Lu. 2010. Contrast and phase combination in binocular vision. *Plos One* 5, 12 (2010), 1–6.
- [24] Frederick A. A. Kingdom and P. W. Woessner. 2021. Binocular summation and efficient coding. *Vision Research* 179 (2021), 53–63.
- [25] D. C. Knill, D. Kersten, and P. Mamassian. 1996. *Implications of a Bayesian Formulation of Visual Information for Processing for Psychophysics*. Cambridge University Press, 239–286.
- [26] Frank L. Kooi and Alexander Toet. 2004. Visual comfort of binocular and 3D displays. *Displays* 25, 2 (2004), 99–108.
- [27] Bernard C. Kress. 2019. Digital optical elements and technologies (EDO19): Applications to AR/VR/MR. In *Digital Optical Technologies 2019*, Vol. 11062. 1106222.
- [28] Chang-Kun Lee, Soon gi Park, Seokil Moon, Jong-Young Hong, and Byoung-ho Lee. 2015. Compact multi-projection 3D display system with light-guide projection. *Optical Express* 23, 22 (2015), 28945–28959.
- [29] Gordon E. Legge and Gary S. Rubin. 1981. Binocular interactions in suprathreshold contrast perception. *Perception & Psychophysics* 30, 1 (1981), 49–61.
- [30] Willem J. M. Levelt. 1965. Binocular brightness averaging and contour information. *British Journal of Psychology* 56, 1 (1965), 1–13.
- [31] Willem J. M. Levelt. 1965. On binocular rivalry. (1965), 1–118.
- [32] Hsin-Hung Li, Marisa Carrasco, and David J. Heeger. 2015. Deconstructing interocular suppression: Attention and divisive normalization. *PLoS Computational Biology* 11, 10 (2015), 1–26.
- [33] Gokhan Malkoc and Frederick A. A. Kingdom. 2012. Dichoptic difference thresholds for chromatic stimuli. *Vision Research* 62 (2012), 75–83.
- [34] Rafal Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. 2006. A perceptual framework for contrast processing of high dynamic range images. *ACM Transactions on Applied Perception* 3, 3 (2006), 286–308.
- [35] Kazumichi Matsumiya, Ian P. Howard, and Hirohiko Kaneko. 2007. Perceived depth in the ‘Sieve Effect’ and exclusive binocular rivalry. *Perception* 36 (2007), 1002–990.
- [36] Rainer Mausfeld, Gunnar Wendt, and Jürgen Golz. 2014. Lustrous material appearances: Internal and external constraints on triggering conditions for binocular lustre. *i-Perception* 5 (2014), 1–19.
- [37] Tim S. Meese, Daniel H. Baker, and Robert J. Summers. 2017. Perception of global image contrast involves transparent spatial filtering and the integration and suppression of local contrasts (not RMS contrast). *Royal Society Open Science* 4, 9 (2017), 170285.
- [38] Yumi Mori, Kohsei Tanahashi, and Satoshi Tsuji. 2000. Quantitative evaluation of visual performance of liquid crystal displays. In *Algorithms and Systems for Optical Information Processing IV*, Vol. 4113. 242–249.
- [39] Robert P. O’Shea and Randolph Blake. 1987. Depth without disparity in random-dot stereograms. *Perception & Psychophysics* 42 (1987), 205–214.
- [40] Robert Patterson, Marc Winterbottom, Byron Pierce, and Robert Fox. 2007. Binocular rivalry and head-worn displays. *Human Factors* 49, 6 (2007), 1083–1096.
- [41] Wolfgang Pieper and Ira Ludwig. 2001. Binocular vision: Rivalry, stereoscopic lustre, and sieve effect. *Perception (Suppl.)* 30 (2001), 1–1.
- [42] Sharon X. Qiu, C. L. Caldwell, J. Y. You, and Janine D. Mendola. 2020. Binocular rivalry from luminance and contrast. *Vision Research* 175 (2020), 41–50.
- [43] Kavitha Ratnam, Robert Konrad, Douglas Lanman, and Marina Zannoli. 2019. Retinal image quality in near-eye pupil-steered systems. *Optics Express* 27, 26 (2019), 38289–38311.
- [44] Jenny C. A. Read and Richard A. Eagle. 2000. Reversed stereo depth and motion direction with anti-correlated stimuli. *Vision Research* 40 (2000), 3345–3358.
- [45] Feng Shao, Weisi Lin, Shanbo Gu, Gangyi Jiang, and Thambipillai Srikanthan. 2013. Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics. *IEEE Transactions on Image Processing* 22, 5 (2013), 1940–1953.

- [46] Jan Skerswetat and Peter J. Bex. 2023. InFoRM (Indicate-Follow-Replay-Me): A novel method to measure perceptual multistability dynamics using continuous data tracking and validated estimates of visual introspection. *Consciousness and Cognition* 107 (2023), 103437.
- [47] Jan Skerswetat, Monika A. Formankiewicz, and Sarah J. Waugh. 2018. Levelt's laws do not predict perception when luminance- and contrast-modulated stimuli compete during binocular rivalry. *Scientific Reports* 8, 1 (2018), 14432.
- [48] Freepik Company 2019. Freepik Mobile App Icon Images. Retrieved April 19, 2019 from <https://www.freepik.com/>
- [49] William W. Sprague, Emily A. Cooper, Ivana Tošić, and Martin S. Banks. 2015. Stereopsis is adaptive for the natural environment. *Science Advances* 1, 4 (2015), e1400254.
- [50] Jonathan Tong, Robert S. Allison, and Laurie M. Wilcox. 2020. Optical distortions in VR bias the perceived slant of moving surfaces. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR '20)*. 73–79.
- [51] Minqi Wang and Emily A. Cooper. 2021. A re-examination of dichoptic tone mapping. *ACM Transactions on Graphics* 40, 2 (2021), Article 13, 15 pages.
- [52] Minqi Wang, Jian Ding, Dennis M. Levi, and Emily A. Cooper. 2022. The effect of spatial structure on binocular contrast perception. *Journal of Vision* 22, 12 (2022), 7–7.
- [53] Gunnar Wendt and Franz Faul. 2019. Differences in stereoscopic luster evoked by static and dynamic stimuli. *i-Perception* 10 (2019), 2041669519846133.
- [54] Gunnar Wendt and Franz Faul. 2022. Binocular luster—a review. *Vision Research* 194 (2022), 108008.
- [55] Xuan S. Yang, Linling Zhang, Tien-Tsin Wong, and Pheng-Ann Heng. 2012. Binocular tone mapping. *ACM Transactions on Graphics (SIGGRAPH Conference Proceedings)* 31, 4 (2012), Article 93, 10 pages.
- [56] S. L. P. Yasakethu, C. T. E. R. Hewage, W. A. C. Fernando, and A. M. Kondo. 2008. Quality analysis for 3D video using 2D video quality models. *IEEE Transactions on Consumer Electronics* 54, 4 (2008), 1969–1976.
- [57] Lili Zhang, Michael J. Murdoch, and Romain Bachy. 2021. Color appearance shift in augmented reality metamerism matching. *Journal of the Optical Society of America A* 38, 5 (2021), 701–710.
- [58] Zhuming Zhang, Chu Han, Shengfeng He, Xueting Liu, Haichao Zhu, Xinghong Hu, and Tien-Tsin Wong. 2019. Deep binocular tone mapping. *The Visual Computer* 35, 6 (2019), 997–1011.
- [59] Zhuming Zhang, Xinghong Hu, Xueting Liu, and Tien-Tsin Wong. 2018. Binocular tone mapping with improved overall contrast and local details. *Computer Graphics Forum* 37, 7 (2018), 433–442.
- [60] Fangcheng Zhong, George Alex Koulieris, George Drettakis, Martin S. Banks, Mathieu Chambe, Frédo Durand, and Rafal K. Mantiuk. 2019. DiCE: Dichoptic contrast enhancement for VR and stereo displays. *ACM Transactions on Graphics (SIGGRAPH Asia Conference Proceedings)* 38, 6 (2019), Article 211, 13 pages.

Received 10 February 2023; revised 14 July 2023; accepted 9 August 2023