Visual Search in Augmented Reality: Effect of Target Cue Type and Location

Amelia C. Warden[1], Christopher D. Wickens[1], Domenick Mifsud[2], Shannon Ourada[2], Benjamin A. Clegg[1], Francisco R. Ortega[2]
Colorado State University Departments of [1]Psychology and [2]Computer Science

The current experiment examined the relative benefit of different cueing aids during a visual target search task, and the tradeoff between reduced information access effort and increased overlay clutter. Using an augmented reality head-mounted display (AR-HMD), participants completed a 180-degree visual search task with three different cue types (world-referenced arrow, screen-referenced icon target image, and screen-referenced minimap) compared to a control condition (no cueing aid). Target cues differed in terms of display proximity and where they were presented on the AR-HMD (the central field of view or approximately 15 degrees downward from the center). We found an overall performance benefit when searching for an object in the far domain with a target cue compared to searching with no cue, and the arrow cue (highest display proximity) showed the greatest overall benefit. We also found a performance benefit for cues located at the center of the AR-HMD compared to the downward location, but this benefit was offset by the higher clutter of the icon image and the minimap. These findings suggest that target cues with higher display proximity that also reduce information access effort (scanning) may be more suitable cueing aids when searching for an object in the far domain.

INTRODUCTION

Visual search is challenging (Drury, 1990) and may be particularly so in outdoor naturalistic environments. For the military, this may often involve searching for an enemy target or weapon in a visually cluttered environment, whether this is a naturalistic terrain or an urban landscape (Wickens & Rose, 2001). It is equally evident that search performance can be greatly aided by cues that designate, by some means, the likely target (Wickens, McCarley, & Gutzwiller, 2022). Of course, the success of such automated cueing depends on the capabilities of an automated agent that can both "know" the target that the searcher seeks and reliably recognize its presence in the environment using some form of image analysis and machine vision.

Computer-supported augmented reality (AR) on a head-mounted display (HMD) stands in a unique position to aid the searcher via such target cueing. Its value is particularly realized because it can visually designate the suspected target while the searcher looks directly at the "search field." This contrasts with looking at an electronic map or target image viewed on a head-down display (e.g., a tablet or mobile device) so that scanning and head movement is required. Furthermore, with reasonable computational power, the system can implement a variety of cueing techniques: for example, a verbal target description, an image of what the target looks like, a map designating the expected location of the target, or even using AR with an arrow pointing to the likely target (or target location) or a reticle surrounding that location. In these AR examples, the position of the target cue on the HMD will move in synchrony with head rotation.

The unique benefit of HMD cueing is manifest in the proximity compatibility principle (Wickens & Carswell 1995, Wickens, McCarley, & Gutzwiller, 2022), which states that:

1. To the extent that two elements in the visual world need to be cognitively integrated (*close mental proximity*) for a task, such integration will be better if they are close together on the display or in the environment (close *perceptual compatibility*).
2. To the extent that they do **not** need to be integrated (more distant mental proximity), they need not be in close spatial proximity, and sometimes task performance in a non-integration task (requiring focused attention on one element) may be better if separated (distant perceptual proximity).

In the context of HMD-supported visual search, the task that relates the cue to its counterpart in the scene beyond is an integration task (close mental proximity). And close perceptual proximity is created by superimposing the two images via the see-through HMD.

Within this context, the proximity compatibility principle has been supported because aircraft HUDs (but not yet HMDs) have been found to be particularly beneficial for the tasks that require integration of display information in the "near domain" (on the HUD) and world information in the "far domain" airspace or terrain below (Fadden, Wickens, & Ververs, 1998; 2000; Wickens, 2021). Furthermore, using "conformal imagery" implemented on the HUD provides an enhanced benefit to integration task performance (Fadden et al., 2000). This is created by the superimposition of the HUD information over the far domain in a way that the position and contours of the image "conform to" those of its far-domain counterpart (e.g., an artificial horizon overlaying the true horizon). The linkage between conformal imagery on the HUD and AR on the HMD is direct, because a property of both is that the imagery *moves in synchrony* with head movement in order to keep a displayed image close to its real-world counterpart as the airplane (HUD) or observer's head

(HMD) rotates. Of the HMD cues described above, the reticle or arrow cue possesses this AR property.

Of the other two spatial cueing examples described above, both the HMD map (a "minimap") and the HMD depiction of the target image are not in AR. However, both do achieve some perceptual proximity with the target because the image is superimposed over the scene (target field) within which the target is embedded, rather than positioned below on a tablet or handheld display.

A series of studies by Yeh and her colleagues (Yeh, Wickens & Seagull, 1999; Yeh & Wickens, 2001; Yeh, Merlo, et al., 2003; Wickens & Rose, 2001) examined HMD target cueing in a military context, using a virtual environment (CAVE) to present the far domain scene. These all revealed the benefit of AR-HMD cueing over similar cues presented with a head-down display. However, they also revealed a smaller cost to the HMD as described in the second part of the proximity compatibility principle explained above. Specifically, the close proximity created by the overlaying imagery disrupted the detection of small uncued targets in the far domain, a manifestation of what is described as "overlay clutter." Fadden et al. (2000, 2001) observed the same phenomenon with the HUD.

The present investigation contained some overlapping features with the studies by Yeh & colleagues reviewed above; however, in the current study, the overall benefits of cueing are examined compared to an uncued condition. Also, in the current study, rather than comparing HMD cueing with a head-down location, all cues were either presented directly in the forward field of view (FFOV) of the HMD or approximately 15 degrees downward in visual space so that scanning and or head movement was required to bring them into foveal vision. Finally, whereas the studies by Yeh and colleagues only examined an AR cue (the reticle or the arrow), the current study compared the effectiveness of the AR arrow cue with two alternatives: (1) the minimap, which cued the location of the target on a top-down translucent image of the entire search field but provided no details of its appearance, and (2) the "icon" cue, which provided an image of what the target looked like, but contained no spatial guidance as to where it was. Thus, we can directly compare the benefits of these two properties of a cue (i.e., appearance and location). The minimap and the icon (image) cues were positioned in the same place on the HMD irrespective of head movements such that both were in screen-referenced coordinates rather than world-referenced coordinates (i.e., neither were AR). This allows for establishing the value of AR by comparing the benefit of these screen referenced cues with that of the AR arrow cue whose position is world referenced. Also, the arrow cue is a far simpler image than the other two, particularly the minimap. Therefore, the arrow cue is less likely to clutter the far domain target field. As noted above, in the cueing conditions, all three were presented either in the center of the field of view, where any clutter might be higher, or displayed downward on the HMD in a location that required some downward head motion to access but would impose less clutter on the forward view. Thus, we can examine the tradeoff between reduced information access effort (IAE; scanning) and increased overlay clutter.

In the current experiment, the search environment was a real (not VR simulated) room, around which were positioned several 3D objects assembled from 3D Mega Blocks, varying in shape and color, one of which was designated as the target at the beginning of each search trial. We predicted that (H_1) cueing would provide an overall benefit compared to the no cue (control) condition. In addition, (H_2) the benefit of cueing would be greatest for the world-referenced AR arrow cue (highest display proximity), less for the screen-referenced icon (lower display proximity), and less still for the screen-referenced minimap because of its greater clutter cost (more complex imagery). Lastly, (H_3) the downward position of the cue, imposing greater scanning requirements to access the cue, will provide a reduced cueing benefit, but this reduced benefit will be offset specifically for the minimap cue because of its greater clutter-reduction of the forward view when moved from the center to the downward location.

METHOD

Participants

A total of 25 participants (7 female, 18 male) volunteered to complete the experiment. Participants consisted of students and staff from Colorado State University, as well as people who were not affiliated with the University. The sample size was limited because of time-constraints and COVID restrictions. All participants had normal or corrected-to-normal vision.

Task

Participants completed the experiment using the HoloLens 2 (AR-HMD). The HoloLens 2, developed by Microsoft Corporation, is a mixed-reality headset that overlays virtual content onto the far domain. The field of view of the device is 43° by 29°.

Participants were seated in a chair at the center of a room and completed a visual search task where they were asked to locate real-world objects (e.g., 3D Mega Blocks) with and without the aid of target cues presented via the AR-HMD (Figure 1). The search field of the room was within 180 degrees around the participant. A total of 32 objects were uniformly distributed across the 180-degree search field in the horizontal direction and approximately 15 degrees in the vertical direction relative to the chair. Objects were placed at the ground, table, and shelf level. The device used the participant's direction of gaze to determine which target object they were looking at in the far domain. When they rested their gaze on an object, a bounding box would appear around the object so the participant knew which one they would be selecting. Participants then used the 'enter' button on a number pad to make their responses.

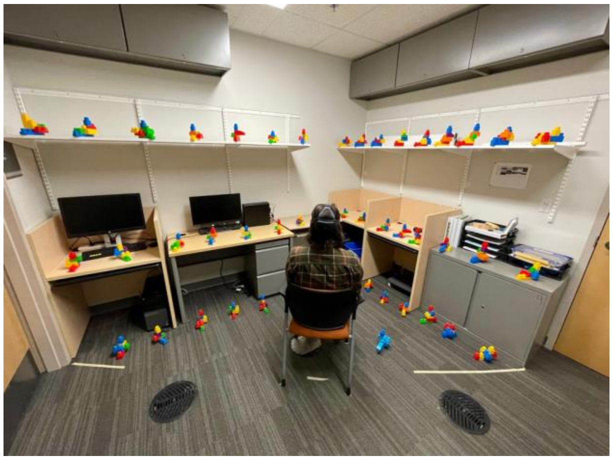

Figure 1. Depiction of the search environment. The objects shown here are slightly more complex than those used in the actual experiment and were used in a later experiment which provided equivalent results.

Shown in Figure 2, the arrow cue was a 2D AR arrow depicted in world-referenced coordinates and pointed in the direction of the target until that target was located. The icon cue was an image of the target object displayed in screen-referenced coordinates and was continuously visible at a fixed location on the HMD as the participant looked for the objects within a 180-degree visual field. The minimap cue, depicted in screen-referenced coordinates, was a 360 degree top-down depiction of the location of all the possible objects in the far domain with the target object highlighted on the map.

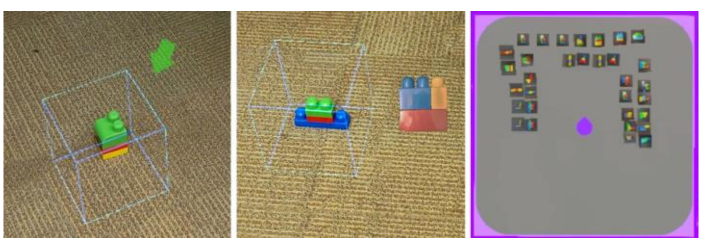

Figure 2. Example images of the arrow, icon, and minimap cue, respectively. The box surrounding the object appears when the participant fixates on the object. In the center image, the icon representation of the target is on the right.

Target cues were presented either at the central field of view (referred to as the 'center' location) or approximately 15 degrees downward from the center (referred to as the 'down' location) on the AR-HMD.

Design

All participants completed eight practice trials (2 for each cueing condition) where they received correct or incorrect feedback. After this, for each target cue (arrow, minimap, icon), participants completed a total of eight test trials. Of the eight test trials for each cue type, four presented the cue in the center location and four in the down location. The cue conditions and location of cue were counterbalanced using the Latin Square design, and trials within each cue condition and cue location were randomized in terms of target location. For the no cue condition, participants completed two sets of 4 trials that were included in the cue condition counterbalancing previously described. The entire experiment consisted of 40 trials and lasted approximately 30 minutes.

Procedure

All participants gave informed consent before they began the experiment. Participants completed a pre-calibration process with the Hololens 2, ensuring accurate eye-gaze detection. Participants read instructions about the visual search task and were told to make their responses as rapidly and accurately as possible. Participants were told to look at a designated mark on the wall at the beginning of the experiment and to return to this location at the beginning of each trial.

RESULTS

Data from one participant was excluded from the analysis because that participant had chance performance. Data were analyzed in R using one-way repeated measures ANOVAs to examine the effect of target cue conditions (arrow, minimap, icon, no cue) on response time and percent error. Next, we conducted a 3 (target cues) x 2 (location) repeated measures ANOVA to examine the effect of location and its interaction with target cue type. All response time data were log-transformed because the data were positively skewed.

Effect of Target Cue Conditions

Response Time. Figure 3 presents the mean response time for each type of target cue, including the no cue condition. The assumption of sphericity was violated for response time data; therefore, we report the p-value of the Greenhouse-Geisser correction for the ANOVA. An ANOVA showed a large effect of cue type on the log-transformed response time, $F(3, 69) = 17.8$, $p < .001$, $\eta_p^2 = 0.44$. Participants located the object faster when using a cue compared to no cue (rightmost bar). Pairwise t-test revealed that participants found the target object significantly faster when using the arrow cue (M = 4.87 s) compared to the no cue condition (M = 10.2 s, $t(23) = -5.52$, $p < .001$, $d = 1.28$), icon cue (M = 9.21 s, $t(23) = -7.18$, $p < .001$, $d = 1.53$), and minimap cue (M = 6.67 s, $t(23) = -2.91$, $p < .008$, $d = 0.53$). They were also faster using the minimap cue compared to the no cue condition ($t(23) = 2.96$, $p = .007$, $d = 0.82$) and to the icon cue ($t(23) = 4.60$, $p < .001$, $d = 0.96$).

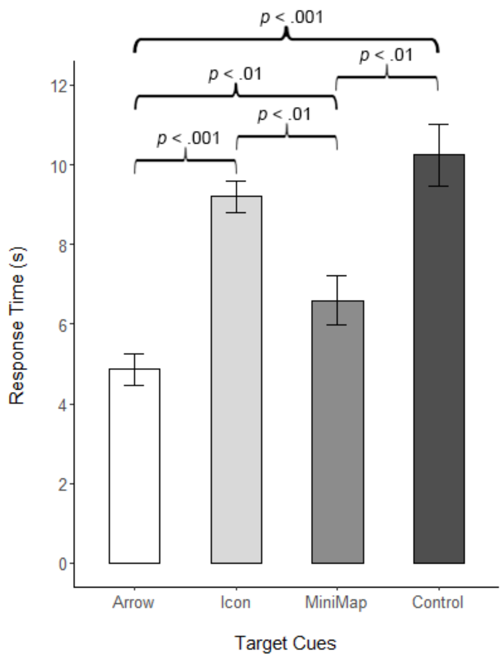

Figure 3. Mean response time plotted as a function of target cue. Error bars represent one standard error of the mean.

Percent Error. Figure 4 presents the mean percent error. An ANOVA showed a significant effect of cue type on F(3, 69) = 2.68, p = .054 (Greenhouse-Geisser correction p = .085), $\eta_p^2 = 0.10$. Pairwise comparisons showed that participants made significantly fewer errors with the arrow cue (M = 7.3%) compared to the no cue condition (M = 18.8%, $t(23) = -2.10$, $p = .048$). There was a marginally significant reduced error rate between the icon cue (M = 10.9%) compared to the no cue condition (M=18% $t(23) = 1.81$, p = .08, $d = 0.50$). The minimap and icon did not differ in accuracy from the control condition.

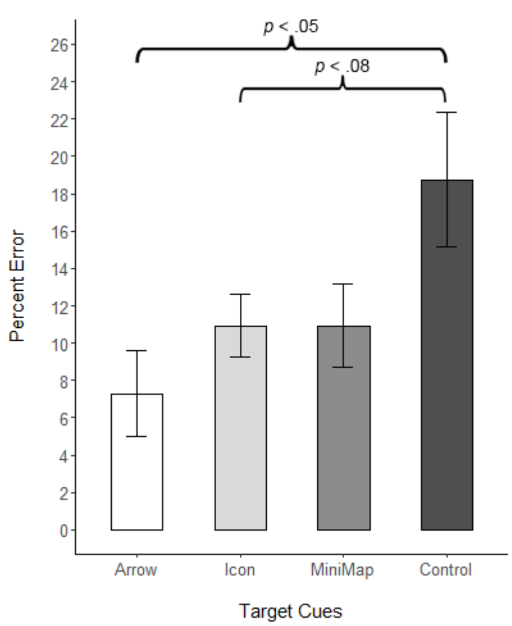

Figure 4. Mean percent error plotted as a function of target cue. Error bars represent one standard error of the mean.

Effect of Target Cue Type and Location

Response Time. An ANOVA on the response time data revealed a large effect of target cue type on the log-transformed response time, F(2, 46) = 28.3, $p < .001$, $\eta_p^2 = 0.55$. The effect of target cue location was also significant, F(1, 23) = 5.44, $p < .03$, $\eta_p^2 = 0.19$, with faster response time when target cues were located in the center (M = 6.48 s) compared to when located downward (M = 7.29 s), $t(23) = 2.33$, $p = 0.03$, $d = 0.35$. A significant interaction between target cue type and target cue location, F(2, 46) = 3.67, p = .03, $\eta_p^2 = 0.14$, indicated a large 2.2 s advantage for the center location for the icon cue , t(23) = 2.93, p = .008, 95% CIs [0.04, 0.24], $d = 0.72$, but not for the arrow cue $t(23) = 1.31$, p = .20, 95% CIs [-0.03, 0.11], $d = 0.23$) and the minimap cue actually showed a non-significant cost for the center location. ($t(23) = -0.63$, p = .54, 95% CIs [-.12, 0.10], $d = -0.14$)

Percent Error. There was no significant main effect of target cue type or location of target cue on percent error. In addition, there was not a significant interaction between target cue type and the location of target cue on percent error.

DISCUSSION

The current experiment examined the influence of information access effort (IAE) and clutter imposed by target cueing aids during a visual search task using an AR-HMD. As predicted (H_1), there was an overall benefit of cueing such that target cues decreased search time by 3.35 seconds and improved accuracy by 9% compared to searching for objects without a cue. We also found partial support for H_2. The world-referenced 2D AR arrow cue showed a clear performance benefit compared to the two screen-referenced cue types, supporting the idea of a greater benefit of overlay for an integration task with AR world-referenced coordinates compared to screen-referenced coordinates (Yeh, Wickens & Seagull, 1999).

Contrary to our predictions (H_2), the minimap cue, despite its greater overlay clutter led to better performance than the icon cue. We interpret this to suggest that the cost of clutter for the visually complex minimap was small, relative to the greater benefit it provided over the icon cue by providing direct guidance as to the **location** of the cue, a greater benefit than providing guidance as to the **appearance** of the cue, by the icon. This would substantially reduce search time, as it did. (Figure 3).

The tradeoff between clutter and IAE posited in H_3 was partially supported and may be specific to the type of target cue. The overall benefit of the center cue locations compared to the downward cue locations was significant, suggesting the center cue imposes less IAE, providing a cueing benefit regardless of the clutter imposed by overlaying information. This benefit was conferred on the smaller, simpler arrow cue suggesting that its clutter costs was reduced in the center condition compared to the greater scanning costs in the bottom location. In contrast, the benefit was not conferred on the more complex image of the icon, and particularly not on the most complex "pixel rich," image of the minimap.

Given that the minimap cue presented the entire object array with the target highlighted, and the arrow cue pointed only to the target, it may be the case that in these two conditions, participants blindly followed the cue without careful examination of whether the cued target in the near domain actually matched the object in the far domain or even the object that they saw at the beginning of the trial.

If a cue is based on AI inference in a cluttered scene, it is always possible that the reliability of such an inference, based on machine vision algorithms, can be less than perfect, and when such imperfection is examined empirically in the HMD (Yeh & Wickens, 2001) or simulated HMD (Mifsud et al., 2022), imperfect cueing has been found to reduce cueing. This may result either because the cue user, knowing the imperfection, takes a longer time to inspect the far domain target carefully (hence increasing response time) or because the cue user automatically follows the cue's guidance and commits an error on those infrequent occasions when the cue is incorrect (hence reducing accuracy). Since we did not include imperfect reliability in the current experiment, we cannot ascertain the extent of such an "automation bias" (Mosier et al., 1998; Parasuraman & Manzey, 2010). Current work using the same paradigm is assessing the influence of target cues with imperfect reliability during the visual search task and suggest an automation bias degrading accuracy when the cue is wrong.

CONCLUSIONS

Overall, we did find the large and expected benefit of HMD target cues relative to no cueing aids. In addition, we found different relative benefits of each type of target cue, with the AR arrow cue, which has the closest (highest) display proximity, showing the greatest benefit overall. Furthermore, identifying a possible target location (the minimap) appeared to have a greater benefit than a visual memory aid of target appearance (the icon). In addition, there was a greater benefit for cues placed in the center of the AR-HMD display compared to the bottom, suggesting that the center imposes less information access effort and provides an increased cueing benefit. However, this benefit was only conferred on the simplest cue, the arrow. While it is possible that participants blindly followed each of locational target cues when locating the object, these results show a clear benefit of cueing relatively to a no cueing situation. Future work will attempt to replicate and extend these findings with a paradigm designed to assess performance during a visual search task using target cues with imperfect reliability.

LIMITATIONS

The current experiment had several limitations. First, the target object was always present within the search field, and target cues that provided location or appearance information always provided information about the location of the correct target object. Future experiments will assess the impact of target cues that are imperfectly reliable when presenting either location or appearance information about objects in the far domain. A second limitation to the experiment was the use of non-naturalistic stimuli (i.e., 3D Mega Blocks). Future experiments will improve external validity by using more ecologically valid stimuli (e.g., finding the nearest exit sign or a fire hydrant). Lastly, the participants of this study included primarily students. Future studies will recruit professionals (e.g., ROTC students or other military participants who may have familiarity with locating target objects in the far domain based on information presented on either a HMD or head-down display).

ACKNOWLEDGEMENTS

This research was supported by the Office of Naval Research under grant numbers N00014-21-1-2949 and N00014-21-1-2580. Dr. Peter Squire was the scientific/technical monitor.

REFERENCES

Drury, C. G. (1990). Visual search in industrial inspection. *Visual search*, 263-276

Fadden, S., Ververs, P. M., & Wickens, C. D. (1998). Costs and Benefits of Head-Up Display Use: A Meta-Analytic Approach. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*(1), 16–20. https://doi.org/10.1177/154193129804200105

Fadden, S., Ververs, P. M., & Wickens, C. D. (2001). Pathway HUDs: Are They Viable? *Human Factors*, *43*(2), 173–193. https://doi.org/10.1518/001872001775900841

Fadden, S., Wickens, C. D., & Ververs, P. (2000). Costs and Benefits of Head up Displays: An Attentional Perspective and a Meta Analysis. SAE Technical Paper 2000-01-5542. https://doi.org/10.4271/2000-01-5542

Mifsud, D., Wickens, C., Rodriguez, R. Maulback, M., Crane, P.& Ortega, F. (submitted) The Effectiveness of Gaze Guidance Lines in supporting JTAC's attention allocation. *Proceedings 2022 Annual Meeting of the Human Factors & Ergonomics Society*. Sage

Mosier, K. L., Dunbar, M., McDonnell, L., Skitka, L. J., Burdick, M., & Rosenblatt, B. (1998). Automation Bias and Errors: Are Teams Better than Individuals? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*(3), 201–205. https://doi.org/10.1177/154193129804200304

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, *52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Wickens, C. (2021). Attention: Theory, Principles, Models and Applications. *International Journal of Human-Computer Interaction*, *37*(5), 403–417. https://doi.org/10.1080/10447318.2021.1874741

Yeh, M., Merlo, J. L., Wickens, C. D., & Brandenburg, D. L. (2003). Head Up versus Head Down: The Costs of Imprecision, Unreliability, and Visual Clutter on Cue Effectiveness for Display Signaling. *Human Factors*, *45*(3), 390–407. https://doi.org/10.1518/hfes.45.3.390.27249

Yeh, M., Wickens, C. D., & Seagull, F. J. (1999). Target Cuing in Visual Search: The Effects of Conformality and Display Location on the Allocation of Visual Attention. *Human Factors*, *41*(4), 524–542. https://doi.org/10.1518/001872099779656752

Yeh, M., & Wickens, C. D. (2001). Display Signaling in Augmented Reality: Effects of Cue Reliability and Image Realism on Attention Allocation and Trust Calibration. *Human Factors*, *43*(3), 355–365. https://doi.org/10.1518/001872001775898269

Wickens, C. D. & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors*, *37*(3), 473-494. https://doi.org/10.1518/001872095779049408

Wickens, C. D., & Rose, P. N. (2001). *Human factors handbook for displays*: *Summary of findings from the Army Research Lab's Advanced Displays & Interactive Displays Federated Laboratory*. Thousand Oaks, CA: Rockwell Scientific Co.

Wickens, C.D., McCarley, J. & Gutzwiller, R. (2022). *Applied Attention Theory* (2nd Ed). Taylor & Francis: Rutledge.