

Quantitative Effects of Overlay Clutter and Information Access Effort: Examining the Scan-Clutter Trade-Off in Displays With Geospatial Maps

Amelia C. Warden¹, Christopher D. Wickens¹, Benjamin A. Clegg², and Francisco R. Ortega³

¹Department of Psychology, Colorado State University

²Department of Psychology, Montana State University

³Department of Computer Science, Colorado State University

Overlaying images from multiple geospatial databases increases clutter and imposes attentional costs by disrupting focusing attention on each database and dividing attention when comparing databases. Costs of overlay clutter may offset the benefits of reduced scanning between two images displayed separately. In two experiments, we examine these attention issues using computational metrics to quantify clutter. We also examine how the scan-clutter trade-off is modified by different levels of clutter, display separation, and task attentional requirements. Participants viewed information from a geographical terrain database and a schematic map database and made judgments that required focusing attention on either database or integrating information across both. In Experiment 1, databases were presented as either overlaid or adjacent displays, and in Experiment 2, as either overlay, adjacent, or more separated displays. Results showed that response time was modulated by the magnitude of clutter, spatial separation, and task type. Results also revealed that clutter costs dominated those of spatial separation, particularly in tasks requiring focused attention. A computational feature congestion metric of clutter effectively predicted performance but could be improved by incorporating an overlay component, which amplified the costs of clutter. The results provide design guidelines for overlay displays (e.g., head-mounted displays) that will minimize the scan-clutter trade-off.

Public Significance Statement

Presenting complex information from multiple databases is a critical design consideration. Such information can be displayed either directly overlaid or separated. For example, various types of information, such as weather or road congestion, can be overlaid directly onto the geographical terrain. Whether this information is best presented as directly overlaid or side by side depends on the amount of effort required to access information and the amount of clutter in the display. Our findings show that the negative impact of overlay clutter is not trivial but can be reduced when carefully selecting the color of overlaid information for devices requiring overlay (e.g., head-up displays or head-mounted displays) or by presenting information side by side when appropriate.

Keywords: information access effort, clutter, visual search, proximity compatibility principle, computational modeling

Display clutter, a pervasive real-world issue, has been extensively studied through human factors research (M. R. Beck et al., 2010; Kaber et al., 2008; Kroft & Wickens, 2002; Moacdieh & Sarter, 2015; Pankok & Kaber, 2018). This problem manifests in diverse scenarios, from Transportation Security Administration agents examining cluttered luggage images to air traffic controllers monitoring complex displays for potential collisions. Display clutter impacts performance during visual search tasks (Moacdieh & Sarter, 2015; Wickens et al.,

2022) and considers factors like the set size, object density, visual complexity, display layout, and target-background or target-distractor similarity (Moacdieh & Sarter, 2015). Notably, the negative influence display clutter imposes on performance has been demonstrated in studies addressing searching through maps or geospatial databases (M. R. Beck et al., 2012; Kroft & Wickens, 2002), complex multielement display panels such as the aircraft cockpit (Lazaro et al., 2021), general information displays (Ognjanovic et al., 2019;

This article was published Online First May 23, 2024.

Amelia C. Warden  <https://orcid.org/0000-0002-8916-1825>

This research was supported by the Office of Naval Research under Grants N00014-21-1-2949 and N00014-21-1-2580 (Francisco R. Ortega). All data, stimulus materials, and analysis scripts are available at <https://osf.io/yujca/>.

Amelia C. Warden played a lead role in conceptualization, data curation, formal analysis, methodology, resources, and software and an equal role in validation, visualization, writing—original draft, and writing—review and editing. Christopher D. Wickens played a lead role in supervision and

conceptualization and an equal role in formal analysis, funding acquisition, writing—original draft, and writing—review and editing. Benjamin A. Clegg played a supporting role in validation and writing—review and editing and an equal role in funding acquisition and supervision. Francisco R. Ortega played a lead role in funding acquisition, a supporting role in writing—review and editing, and an equal role in supervision.

Correspondence concerning this article should be addressed to Amelia C. Warden, Department of Psychology, Colorado State University, 1876 Campus Delivery, Fort Collins, CO 80523-1876, United States. Email: amelia.warden@colostate.edu

Pankok & Kaber, 2018), and see-through displays that overlay a natural scene, such as head-up displays (HUDs; S. H. Kim et al., 2011; Ververs & Wickens, 1998) and head-mounted displays (HMDs; Dey et al., 2018; Warden, Wickens, Clegg, & Ortega, 2023).

The current experiments model the scan-clutter trade-off by examining the effects of display clutter and visual scanning on visual search performance for different display configurations. These display configurations consist of two domains of information: (a) the near domain, which is imagery that overlays the scene information, and (b) the far domain, which is the scene imagery that is perceptually behind the overlaid near-domain imagery. We compare an overlay display, where the two domains of information are overlaid onto the same display or spatial region, to a separated display, where they are presented either side by side or further separated. The overlay display increases the overall display clutter in the scene due to what we term created by superimposing the near domain onto the far domain. Separated displays can mitigate the effects of overlay clutter but can impose visual scanning costs between spatially separated domains (Ballard et al., 1995; Draschkow et al., 2021; Gray & Fu, 2004; Lu et al., 2020; Wickens, 2014; Yang et al., 2014). This cost is referred to as *information access effort* (IAE). By modeling the scan-clutter trade-off for visual search tasks requiring different attentional demands, we can predict the quantitative effects of clutter.

Clutter: Causes, Modeling, and Quantification

The term “clutter” is multifaceted. Rosenholtz et al. (2007) describes clutter as an excess of items, or their representation or organization, that degrade task performance. Moacdieh and Sarter (2015) referred to “display clutter” as the performance and attentional costs due to the interaction between high data density, poor organization, and irrelevant information. Kaber et al. (2008) defined clutter as the “unintended effect of display imagery that obscures or confuses other information or that may not be relevant to the task at hand.” Collectively, clutter is characterized as a crowded or unorganized arrangement of information that can obscure or confuse information and impose performance and attention costs.

Clutter presents challenges in applied visual search tasks (Wickens et al., 2022; Wolfe, 2021). Excessive clutter prolongs the target identification time, which is crucial in time-sensitive scenarios like airport baggage screening or navigation with HUD directional information while driving. Clutter may even prevent target detection (Wolfe & Horowitz, 2004), which can have disastrous consequences for safety-critical tasks, like a soldier on the battlefield using an HMD to guide attention to a hostile target. Given that clutter can compromise both the speed and accuracy of visual search in safety-critical, real-world tasks (Moacdieh & Sarter, 2015), the human factors community needs to model and predict its effects on both search time and accuracy. Despite the reduced coverage of clutter in basic vision research (e.g., see Donderi, 2006; Wolfe, 2021), six fundamental concepts and their computational models from basic visual attention can be used to identify different factors of clutter and to evaluate the effectiveness of these models when predicting search time and estimating the amount of clutter in real-world displays and scenes.

Numerosity clutter is a source of clutter related to the number of discrete objects in the field of view, where an increase in objects increases the visual complexity of the display (Yeh & Wickens, 2001). In visual search tasks, the number of distractors, indicated as N or numerosity, directly impacts performance, particularly during

serial tasks (Wolfe, 2021), where search time increases linearly with N . Numerosity clutter is easily quantified by an object count based on the individual alphanumeric characters or geometric objects in the search field. Numerosity clutter has been effective at predicting search time in clutter studies examining human-generated maps, consisting of features like mountains defined by contour lines, symbols, and roads (M. R. Beck et al., 2012; Yeh & Wickens, 2001). However, Neider and Zelinsky (2011) highlighted the challenge of defining objects when quantifying clutter in naturalistic scenes. For instance, when viewing a forest scene, what constitutes a single or multiple object(s) is subjective compared to human-rendered maps, which objectively categorize distinct representations of objects. In the current research, we systematically manipulate the number of objects in the near-domain display imagery into three distinct levels of numerosity clutter (low, medium, and high) based on an object count metric.

Clutter can also be attributed to spatial proximity and global density clutter (GDC). Close spatial proximity of relevant and irrelevant objects impairs focused attention on the relevant objects (Broadbent, 1982; Eriksen & Eriksen, 1974). The attentional “spotlight” is approximately 1° of visual angle wide (Broadbent, 1982); therefore, when irrelevant items fall within that attentional “spotlight” for tasks demanding focused attention, they interfere with performance (Eriksen & Eriksen, 1974). Close spatial proximity has been used to quantify clutter (Tullis, 1983, 1986) by exploring the costs of densely packed versus more separated elements in a display, referred to as GDC (see Wickens et al., 1997, for a review). GDC can vary independently of numerosity clutter by tightly packing the same number of search elements in a smaller area. Regarding density clutter, the edge density clutter metric (Neider & Zelinsky, 2011; Rosenholtz et al., 2007) quantifies clutter based on the “edges” and “density” of objects in a display. While this metric predicts clutter effects on search time (Henderson et al., 2009; Neider & Zelinsky, 2011), it fails to capture color variability and is highly sensitive to parameter settings, like thresholds for determining individual edges (Rosenholtz et al., 2007). In the current research, overlaying near-domain display imagery onto the far-domain scene imagery increases the density of clutter between elements in both domains and increases the spatial proximity between individual elements.

While GDC applies to the entire display (i.e., independent of the target in a search task), local density clutter (LDC) or crowding occurs when nonrelevant items closely surround a single target element to be found or identified. LDC has a greater influence on performance when GDC is high (M. R. Beck et al., 2012). From this, we predict that the influence of LDC will be greater when GDC is high (a statistical interaction), such as when overlaying the near- and far-domain imagery compared to presenting them separately.

Key elements for search and clutter modeling are object attributes and feature heterogeneity, or the degree of variability across feature types. According to the feature integration theory (Andersen et al., 2021; A. M. Treisman & Gelade, 1980; A. Treisman & Paterson, 1984), objects are defined by discrete features (e.g., shape, size, color) with different degrees of variability in a search field. Greater feature variance of nontarget items inhibits search performance (Duncan & Humphreys, 1989). For instance, a target distinguished by a single feature (e.g., color) can be found quickly via parallel search and is minimally impacted by the number of irrelevant objects in the search field (Wolfe, 2021). However, targets formed by two features (e.g., shape and color) in displays with more

irrelevant features defining objects result in a slow, serial, and less efficient search because each item requires inspection (Wolfe, 2021).

Two computational clutter metrics have exploited these findings to predict clutter effects. The “color cluster clutter” (C3) metric focuses on the single object attribute of color and spatial proximity of objects to predict both the subjective perception of clutter (Lohrenz et al., 2009) and objective search time in map displays (M. R. Beck et al., 2012). The results revealed an interaction between LDC and GDC (M. R. Beck et al., 2012). Only low color variance (i.e., high similarity) between the target and background led to slower searches as global clutter increased, especially in high local clutter areas. Irrelevant objects compete for attention as a function of clutter in low-color-variance displays.

The feature congestion metric assesses feature heterogeneity across all features in a display and penalizes displays with higher feature variability in the search field (Rosenholtz et al., 2007). In highly cluttered displays, feature space is already congested, so adding more features fails to capture attention. In the context of color, high color variability represents more clutter and increases search time (Lohrenz et al., 2009; Rosenholtz et al., 2007). The feature congestion metric has been shown to predict response time for target search in such naturalistic scenes (Henderson et al., 2009; Neider & Zelinsky, 2011), where defining “what an object is” becomes difficult (see also Henderson et al., 2009).

In the current research, we used the feature congestion metric (Rosenholtz et al., 2005, 2007) to measure visual clutter because it best captured empirically tested aspects of clutter by considering several perceptual features (e.g., predicting contrast thresholds, texture, and orientation clutter) and was just as effective in predicting performance effects as the subband entropy and edge density metrics (see Rosenholtz et al., 2007). Furthermore, unlike other metrics, the feature congestion better captures the influence of color variability on visual clutter (Rosenholtz et al., 2007). Therefore, the feature congestion metric appeared to be the most comprehensive.

The above models penalize clutter for greater heterogeneity or variance, but the present work addresses an opposing effect of clutter often not included in clutter models: greater similarity between the search target and distractors. We specifically focus on the impact color has on quantifying clutter. Prior research shows that searching for a single target among distractors is easier when the target differs in color (Bundesen & Pedersen, 1983; Holahan et al., 1978). Color dissimilarity creates a pop-out effect, aiding focused attention tasks. Conversely, when integration is required, color similarity facilitates performance, whereas color dissimilarity does not (Wickens & Andre, 1990). In one study (Nagy & Sanchez, 1990), small color differences (similar hues of the same color) increased search time with display density (clutter), indicating a serial search. However, the search time was shorter and constant, with large color differences (very different hues of the same color) regardless of the number of distractors, suggesting a parallel search. Overall, the number of items in a display matters most when color differences are small, with color similarity differentially impacting tasks with different attentional demands.

Last, organization is another factor that is known to influence clutter and, consequently, performance (Moacdieh & Sarter, 2015; Tullis, 1983, 1986). However, in the current research, we do not focus on this aspect of clutter, and, therefore, further elaboration is beyond the scope of this article.

The current research applies five basic vision and attention concepts to quantify clutter and predict performance and attention costs of clutter: (a) numerosity clutter; (b) spatial proximity, density clutter, and edge count; (c) LDC; (d) object attributes and feature heterogeneity; and (e) color. The current experiments also address three additional issues of clutter that impact performance and are less explored in prior research. Specifically, issues of overlay clutter, the attentional demands of the task, and when information from one source is spatially separate from a second source. These three issues are described in detail in the next section.

Modeling Overlay Clutter

A less examined clutter cost is the simultaneous consideration of the quantitative influences of global density and local density computational clutter metrics in scenarios that involve viewing a scene through a translucent display on either a HUD (Weintruab & Ensing, 1992; Wickens et al., 2004) or an HMD (Dey et al., 2018; Yeh et al., 1999). While the unwanted effects of such “see-through clutter” have been examined, particularly concerning aircraft HUDs (Fadden et al., 2001; Lee et al., in press; Ververs & Wickens, 1998; Weintruab & Ensing, 1992; Wickens et al., 2004), prior work has not focused on modeling the effects of overlay clutter on either the HUD or HMD platform.

In the current context, overlay clutter closely relates to local density clutter (LDC) by quantifying the cost of overlaying near- and far-domain information for tasks that require locating, interpreting, or identifying information. Two examples where the amount of clutter in specific regions of the display increases include overlaying weather data on a geographical map or HMD information overlaid onto the real-world scene. In the context of overlay clutter, it is important to distinguish between the clutter in the near- and far-domain imagery. The near-domain display imagery can inhibit attention to information beyond the display (e.g., a map presented on an HMD), whereas the far-domain imagery in the naturalistic scene behind the displayed information can inhibit attention to the displayed information rendered closer to the observer (i.e., near-domain imagery).

In the present research, we refer to display clutter in the near domain as one source of overlay clutter that is distinct from scene clutter in the far domain, which is another source of clutter. Both sources can be quantified by computational metrics of clutter (e.g., numerosity and feature congestion), and it is our interest to examine how these quantitative influences are modulated in overlay displays. We examine how these metrics apply to overlay displays by simulating a scenario akin to using an HMD or HUD, which overlays near-domain information onto the far-domain scene to create overlay clutter.

The Attentional Task

A second issue focuses on the attentional task required when comparing information that is overlaid versus separated (e.g., an HMD vs. a handheld tablet). Focused attention tasks require the examination of a single source of information in either the near-domain display or the far-domain scene. For instance, a pilot searching for a mountain and judging its relative height has to filter out all irrelevant map information, such as weather information and roadways.

A second class of attention tasks are those that require dividing attention to integrate information between elements in both domains

(Kroft & Wickens, 2002; Wickens & Carswell, 1995; Wickens & Ward, 2017). For instance, a pilot looking for weather over a specific geographical region to determine if a course change will avoid both terrain and bad weather must integrate weather information relative to geographical information. Here, both domains are the focus of attention, and divided attention is required to integrate information (Wickens et al., 2022). The distinction between focused attention and divided attention integration tasks is important when evaluating clutter effects because of the proximity compatibility principle (Wickens & Carswell, 1995), which predicts that close spatial or perceptual proximity created by overlay displays disrupts focused attention tasks more than integration tasks (Kroft & Wickens, 2002; Wickens & Ward, 2017).

The Control Condition: Display Separation

Critical to our research is the control condition to which an overlay display is compared; specifically, when the near- and far-domain imagery, containing the same information as the overlay display, are viewed separately. For a desktop display presenting multiple sources of information overlaid onto the same plane, the separated display condition could display each layer of information on a separate monitor. For the HUD or an HMD, the separated display condition could be the head-down center console or a handheld display (Yeh et al., 2003). The spatial separation of information will, by definition, eliminate overlay clutter. But spatial separation imposes a different penalty: specifically, visual scanning or IAE, as the eyes must move between the two sources of information, such as the display(s) and scene (Ballard et al., 1995; Draschcow et al., 2021; Gray & Fu, 2004; Lu et al., 2020; Wickens, 2014; Yang et al., 2014).

The performance cost due to visual scanning, sometimes supported by head rotation, increases as spatial separation increases (Poole et al., 2023; Yang et al., 2014). Based on prior research, this IAE penalty should be amplified to the extent that information integration is required between the two domains, as that information in the first-scanned source needs to be maintained in working memory while the second source is located (Vincow & Wickens, 1993). This extra penalty for spatial separation imposed on integration tasks but not focused attention tasks is a fundamental element of the proximity compatibility principle (Wickens & Carswell, 1995) and moderates the trade-off between spatial separation and overlay.

This trade-off is referred to as the scan-clutter trade-off, where spatial separation will disproportionately penalize performance on the integration task due to increased scanning between sources of information. Conversely, overlay will disproportionately penalize performance on the focused attention tasks due to costs associated with overlay clutter. Existing research from meta-analyses on both HUDs (Fadden et al., 2000) and HMDs (Warden, Wickens, Clegg, & Ortega, 2023) suggests the costs of spatial separation exceed the costs of overlay clutter. But the amount of display/scene clutter and the type of attentional demands modulate the benefit of overlay displays.

For computational modeling, just as we apply metrics to clutter effects, we can also turn to computational models of IAE penalties associated with both eye movements and head movements (Warden et al., 2022; Wickens, 1993). The penalties associated with scanning are small; those associated with head movements, particularly larger ones, are greater (Poole et al., 2023), an issue that is examined in more detail in Experiment 2.

Aims and Hypotheses

In the current set of experiments, we examine many of the aforementioned issues, specifically focusing on modeling overlay clutter, attentional tasks, and spatial separation. None of these issues appear to have been coupled with computational clutter modeling by previous researchers. In both experiments, we devised a task that allowed us to manipulate the amount of clutter in the near-domain display and far-domain scene. Participants answered questions regarding elements in both a 3D hybrid map (i.e., consisting of normal and satellite views) of the terrain viewed from above representing the far-domain scene and an electronic 2D map containing text, symbols, and lines to represent additional information (e.g., weather and flight paths) in the near-domain display. Clutter varied in each domain by categorizing levels of clutter as low, medium, and high based on Rosenholtz's (Rosenholtz et al., 2007) feature congestion metric of clutter and, for the near-domain display, by our subjective numerosity object count quantification of display clutter (also referred to as numerosity clutter). The questions required focused attention to one domain or integration across both domains.

Experiment 1 compared overlay and adjacent geospatial maps to examine the scan-clutter trade-off and determine how this trade-off is influenced by the level of computational image clutter, the presence of overlay clutter, and task type. Experiment 2 replicated Experiment 1 except that, in addition to the overlay condition, we compare two separated-display conditions: the original adjacent condition and one separated by a larger distance. Comparison of performance between the adjacent and the separated conditions disentangles the effects of IAE from those of overlay clutter.

In the following hypotheses, the term "performance" refers equivalently to speed (response time) and accuracy (error rate), and "focused attention" applies equally to that about elements in the far-domain scene and those on the near-domain display.

Hypothesis 1: Overlay displays will increase response time relative to adjacent displays because of the well validated impact of overlay clutter on visual search tasks.

Hypothesis 2: An interaction is predicted such that performance for the integration task will benefit from overlay displays more than the focused attention task, where overlay costs may be observed at higher levels of image clutter. This is based on the general predictions of the proximity compatibility principle and the findings of both Kroft and Wickens (2002) and Wickens and Ward (2017).

Hypothesis 3A: Increasing computational clutter on both the near-domain display and in the far-domain scene will impose a cost to the focused attention task, as revealed by most of the search task research reviewed above. We also predict a cost to performance for integration tasks, but we are unsure of the extent to which performance may be reduced by increasing clutter.

Hypothesis 3B: The negative effects of increasing clutter will be amplified when overlay clutter is present and diminished in the separated condition. This predicted interaction between sources of clutter is based on findings of similar interactions by M. R. Beck et al. (2012).

Hypothesis 3C: There will be performance asymmetries as a function of the clutter domain and question domain. Near-

domain display clutter will hurt focused attention on the far-domain scene more than far-domain scene clutter hurts focused attention on the near-domain display. We predict this based on what we refer to as the “biological nearness effect,” which is based on literature showing that behavioral responses increase when information is closer in proximity to the body (see Bufacchi & Iannetti, 2018).

Hypothesis 4: Greater color similarity between the near-domain display and far-domain scene will impose performance costs on both attention tasks. This effect will increase with increasing display clutter.

Hypothesis 5: The two computational metrics of clutter, numerosity clutter as quantified by object count and feature congestion, will predict task performance.

Hypothesis 6 (Experiment 2 only): In Experiment 2, increasing IAE from adjacent to separate will hurt performance, and more so for the integration task than for the focused attention tasks.

Experiment 1

We examined the cost of clutter for two types of map tasks presented on a flat-panel display. Participants viewed a series of maps containing information in the far-domain scene and near-domain display. The amount of clutter in each domain was determined by (a) the feature congestion metric (Rosenholtz et al., 2007) and (b) numerosity clutter (an object count metric for the near-domain display). The near- and far-domain imagery were either overlaid onto the same region in space or presented adjacent to each other (e.g., side by side). Participants made judgments based on information in either one domain (focused attention questions) or both domains (integration questions).

Method

Transparency and Openness

All data, stimulus materials, and analysis scripts are available at <https://osf.io/yujca/>. This study was not preregistered.

Participants

Thirty-six students enrolled in an introductory psychology course at Colorado State University received course credit for completing the experiment. A sample size of 36 would be needed to achieve 80% power with a medium effect size of Cohens $f = 0.25$ ($\alpha = .05$, repeated-measures analysis of variance [ANOVA]). All participants had self-reported normal or corrected-to-normal vision and were screened for colorblindness using an electronic version of Ishihara’s test. Previous work suggests online colorblindness tests are valid measures for screening for general color deficiencies (Rashid et al., 2021; Van Staden et al., 2018).

Stimuli and Apparatus

All stimuli were presented on HP Compaq LA2306 wide monitors measuring 23” diagonally and with a screen resolution of 1,920 × 1,080 pixels. The experiment was created and presented in E-Prime 3.0 software (Psychology Software Tools, 2020,

Pittsburgh, Pennsylvania). The stimuli were created using CalTopo (CalTopo, 2021), an online application used to create topographic maps based on geospatial databases. The far-domain scene imagery was composed of hybrid maps of various geographical regions within the United States and included a mountain icon with the name and elevation data and contour lines with elevation data. The near-domain display imagery consisted of aircraft icons with an aircraft identification tag and altitude information, flight paths indicating the direction of the aircraft, and weather icons (e.g., warm and cold fronts, rain, wind). On different trials, aircraft and flight path icons in the near-domain imagery were either green or red to examine the possible role of color confusion in overlay trials. The near-domain display was overlaid onto the far-domain scene for the overlay display condition and appeared to the right of the far-domain scene for the adjacent display condition, such that the center of the two images was separated by approximately 10°. Images for each domain were 1,000 × 1,000 pixels. Figure 1 provides three illustrations of the near- and far-domain imagery for each map condition.

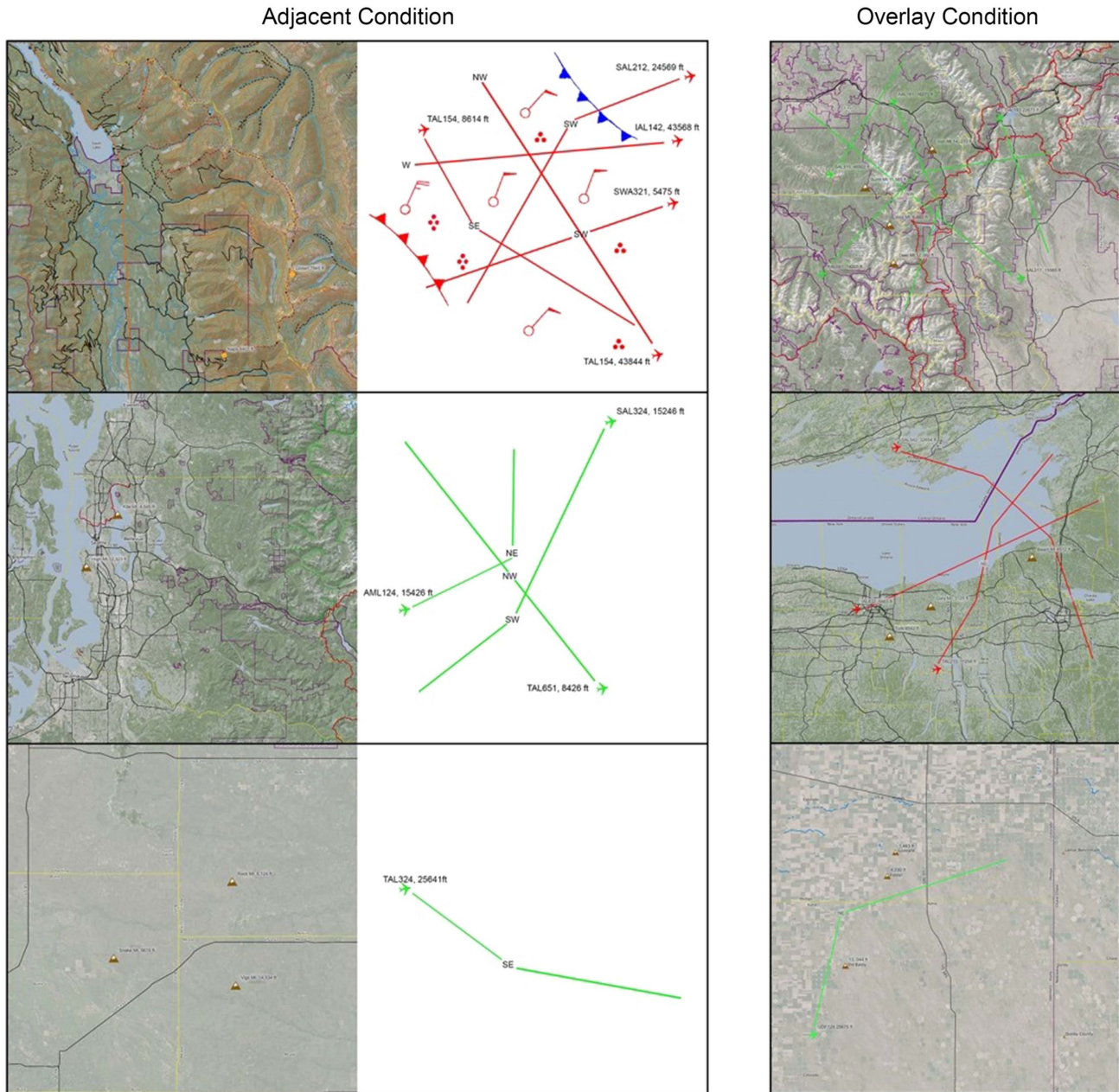
Using Rosenholtz et al.’s (2007) computational clutter model, we verified the magnitude of the feature congestion clutter metric for each domain and the total feature congestion value for the overlay display condition. These values were derived from the three major near-domain clutter categories (low, medium, and high), each of which was replicated with six different images. These six images did not have precisely identical feature congestion values, although the six within each of the three levels of clutter were close to each other. A total of nine unique combinations of near- and far-domain clutter were created using the three levels of clutter (low, medium, and high). Feature congestion values for each domain as well as the overlay display condition are shown in Table 1.

In addition to feature congestion, we employed two other metrics. First, the computational version of numerosity clutter was used to create three levels of near-domain display clutter. Numerosity clutter here is closely related to edge density, in that an object contains one (e.g., a line) or multiple edges. The number of objects in the near domain at the three clutter levels was 9, 16, and 29. As noted above, object count cannot be applied to far-domain scene clutter. Second, overlay clutter was represented as a simple dichotomous variable, either present (overlay condition) or absent (adjacent and, in Experiment 2, the separated condition).

To create the overlay condition, we generated three levels of clutter (low, medium, and high) using an object count quantification of clutter in a near-domain display simulating a 2D electronic map with air traffic and weather information. The feature congestion metric values of these three levels can be seen in the left column of Table 1. We then overlaid these on three different geographical terrains or scene images with three different levels of far-domain scene clutter. The terrain images were rendered as if the terrain were viewed from above by a pilot or a drone. For the far-domain scene, we selected three types of terrains that perceptually represented low, medium, and high clutter based on the number of roads, town names, mountains, and contour lines (see Figure 1 and the center column of Table 1). The joint effects of both near-domain display and far-domain scene clutter on these nine images were assessed via the computational feature congestion algorithm when they were overlaid, thereby including any influence of inducing overlay clutter (right column), and again when they were presented adjacent to or separated from each other by a specific visual angle.

Figure 1

An Illustration of the Stimulus Images for the Adjacent (Left) and Overlay (Right) Display Conditions



Note. The near- and far-domain imagery were separated by 10° of visual angle in the adjacent condition. The top images represent high near- and far-domain clutter, middle images represent medium near- and far-domain clutter, and bottom images represent low near- and far-domain clutter. Trials could feature any combination of near- and far-domain clutter. Near-domain imagery could be either red or green air route and aircraft information. Air route directions are abbreviated using the cardinal directions of North (N), South (S), East (E), and West (W). Aircraft identifiers (e.g., TAL, SWA) represent unique aircrafts. See the online article for the color version of this figure.

Unique questions were created for each unique image. For the integration task, each question required integrating information in the far and near domains. For example, “Is the altitude of aircraft UAL232 higher than the elevation of Ellis Mountain?” In this example, participants must compare the aircraft altitude (near-domain display) to the mountain elevation (far-domain scene).

For the focused attention task, each question required focusing on either the far or the near domain. For example, “Is the flight path of aircraft UAL232 heading NW?” (near-domain display) or “Is the elevation of Ellis Mountain above 12,000 ft?” (far-domain scene). On each trial, the imagery and question appeared simultaneously.

Table 1
Average Feature Congestion (FC) Metric Values for Clutter Levels in Each Domain

| Near-domain clutter level | Near-domain FC value | Far-domain clutter level | Far-domain FC value | Overlay display condition FC value |
|---------------------------|----------------------|--------------------------|---------------------|------------------------------------|
| High | 3.27 | High | 5.86 | 5.43 |
| High | 3.29 | Medium | 4.42 | 4.46 |
| High | 3.29 | Low | 3.08 | 3.33 |
| Medium | 2.09 | High | 5.8 | 5.12 |
| Medium | 2.13 | Medium | 4.43 | 4.08 |
| Medium | 2.12 | Low | 3.15 | 3.03 |
| Low | 1.56 | High | 5.76 | 4.82 |
| Low | 1.58 | Medium | 4.4 | 3.9 |
| Low | 1.55 | Low | 3.13 | 2.89 |

Note. Average feature congestion (FC) values from the Rosenholtz computational clutter model for each domain separately (near and far) and for the overlay condition collapsed across each domain to represent each unique combination of near- and far-domain clutter.

Design

Participants completed four blocks. Blocks were counterbalanced based on task type (focused attention and integration) and display type (overlaid and adjacent). Each block consisted of 36 randomized experimental trials. Participants completed a total of 144 trials plus six practice trials, with auditory feedback for each task indicating whether they were correct or not. There were four trials for each clutter combination. For example, there were four trials for the low far-domain scene and the low near-domain display clutter image combination. To examine the effect of color similarity, air route, and aircraft icons in the near domain were green for half of the trials and red for the remaining half. The correct answer for half of the experimental trials was yes, and the other half was no. For the adjacent condition, half of the trials asked questions about the near domain, and the remaining half asked questions about the far domain. For the integration task, questions pertained to both domains. Participants were exposed to the same stimuli presented in a random order for each block. The entire experiment lasted approximately 30 min.

Procedure

Participants gave informed consent to participate after reviewing the consent documentation presented on a desktop computer. Before the experiment, participants completed an electronic colorblindness test to confirm that they did not exhibit signs of red-green colorblindness. Participants were instructed to maintain proper posture and avoid leaning backward or forward during the experiment. Before the main experiment, participants were given instructions about each task and were trained to read the images and icons. After they completed the training session, they completed the main experiment, including two rounds of practice trials for each type of task. Participants were instructed to answer questions as accurately and rapidly as possible. To make their response, they pressed the “Y” button on the keyboard for “yes” and the “N” button for “no.”

Results

The experimental design consisted of three independent variables: display (overlay, adjacent), task (focused attention near domain,

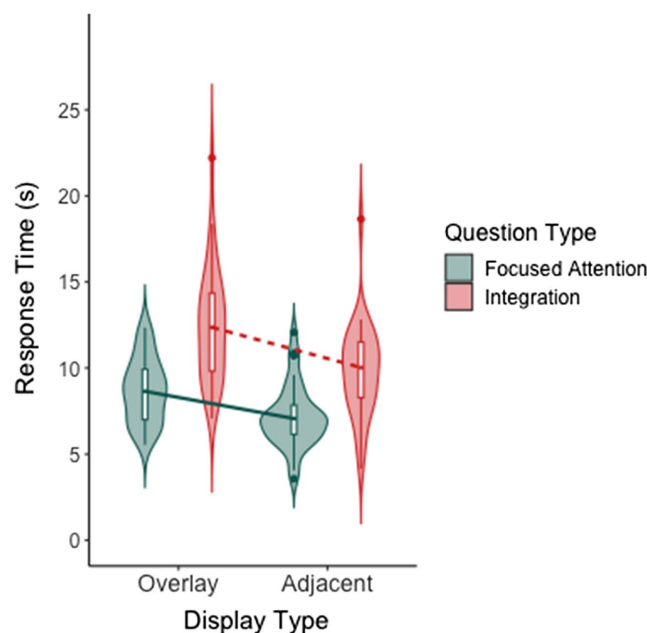
focused attention far domain, and integration), and clutter level. Computational clutter level was integrated into a single nine-level variable representing each combination of near- and far-domain clutter, where each domain had three levels (low, medium, and high; see Figure 1 and Table 1). Our analyses examined the effects of overlay clutter in the focused attention questions separate from those in integration questions. Given that prior research examining the effect of clutter on search has focused on response time rather than accuracy (Moacdieh & Sarter, 2015), we include more details on response time for our computational clutter analysis.

Before conducting any analyses, we looked for outliers using the Grubbs and Rosners tests and boxplots for both response time and accuracy. No outliers were revealed. All data were included in all the analyses described below. All data were analyzed in R.

Effect of Display Separation: IAE Versus Overlay Clutter

We used a two-way repeated-measures ANOVA, collapsed across computational clutter levels, to explore performance differences with spatial separation (overlaid vs. adjacent) for both focused attention and integration tasks. Response time met normality assumptions (Shapiro test and Q-Q plots $p > .05$) for all 2 (display type) \times 2 (question type) conditions. While percent error violated normality assumptions, data were transformed using the square root function. The untransformed data essentially yielded the same effects. Figures 2 and 3 illustrate the impact of task type on response time and percent error, respectively. Untransformed percent error data were plotted for visual clarity.

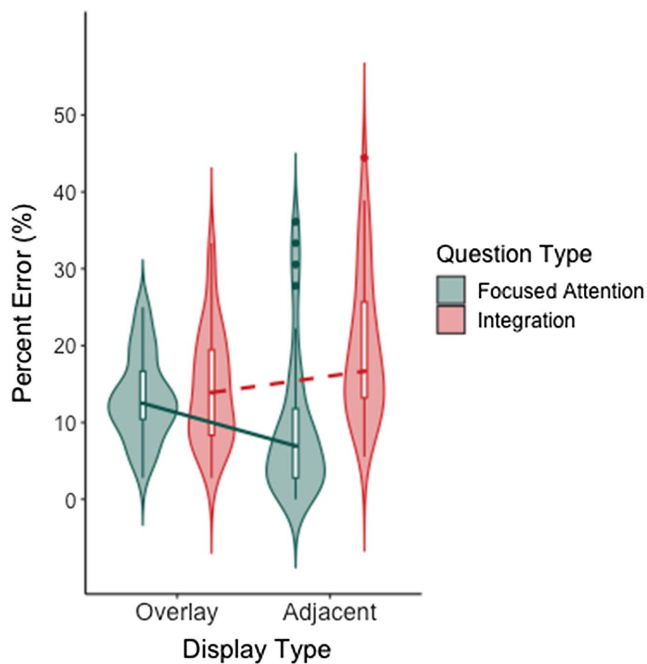
Figure 2
Experiment 1: Mean Response Time Plotted as a Function of Display Type and Task Type



Note. Solid green and dashed red lines represent focused attention and integration task, respectively. Error bars represent one standard error of the mean. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Figure 3
Experiment 1: Mean Percent Error Plotted as a Function of Display Type and Task Type



Note. Solid green and dashed red lines represent focused attention and integration questions, respectively. Error bars represent one standard error of the mean. See the online article for the color version of this figure.

In Figure 2, the analysis of response time revealed a statistically significant main effect of the display, $F(1, 35) = 63.2, p < .001, \eta_p^2 = 0.64$. Supporting Hypothesis 1, participants were, on average, 2.1 seconds (s) slower when answering questions for overlay displays compared to adjacent displays. Additionally, there was a statistically significant main effect of task, $F(1, 35) = 85.62, p < .001, \eta_p^2 = 0.71$. On average, participants were 3.2 s faster when answering questions requiring focused attention compared to integration. The interaction between display and task was statistically significant, $F(1, 35) = 7.03, p = .012, \eta_p^2 = 0.17$. Specifically, there was a greater cost to response time when information was overlaid for the integration task than for the focused attention tasks. Pairwise comparisons revealed a cost of overlay displays for integration questions, $t(35) = 6.98, p < .001, \text{Cohen's } d = 0.88$, and focused attention questions, $t(35) = 5.59, p < .001, \text{Cohen's } d = 0.84$. Contrary to Hypothesis 2, this interaction indicates that the cost of overlay was slightly amplified for the integration task compared to the focused attention tasks.

For the transformed percent error in Figure 3, there was no statistically significant effect of display, $F(1, 35) = 0.62, p = .43, \eta_p^2 = 0.02$. However, there was a statistically significant main effect of task, $F(1, 35) = 79.76, p < .001, \eta_p^2 = 0.68$. As with response time, there was a greater cost to performance for the integration task, which imposed greater demands on working memory than the focused attention task. There was a statistically significant interaction between display and task, $F(1, 35) = 30.79, p < .001, \eta_p^2 = 0.47$, supporting Hypothesis 2 for accuracy. For the overlay

display, pairwise comparisons revealed lower error for the integration task, $t(35) = -4.22, p < .001, \text{Cohen's } d = 0.64$, and, for the adjacent display, lower error for focused attention tasks, $t(35) = 3.85, p = .0004, \text{Cohen's } d = 0.72$.

Effects of Feature Congestion Clutter on Focused Attention Questions

We used linear mixed models (LMMs) and binomial generalized linear mixed models (GLMMs) to examine Hypothesis 3, focusing on the effects of near- and far-domain feature congestion clutter on response time and error rate data. Additionally, we examine asymmetries in performance based on the clutter and question domain. To address the nonindependence present in the data when participants provide multiple responses, participant was included as a random effect in all models with random intercepts. Cohen's d effect sizes for LMMs were computed by dividing the expected mean difference by the square root of the expected pooled variance of an individual observation, where the denominator represents the variation within each condition across participants and stimuli (Westfall et al., 2014).

In the LMM for focused attention questions, the dependent factor was response time (seconds), and the fixed effects were either near- or far-domain feature congestion clutter, question type (near vs. far domain focused attention questions), and their interaction. For all models, question type was submitted as a factor, and participant was included as a random effect with random intercepts. Note that the primary focus of our results is on response time because this measure is the most examined in prior clutter modeling research. Therefore, we only present graphs for response time and include a table for the error analyses.

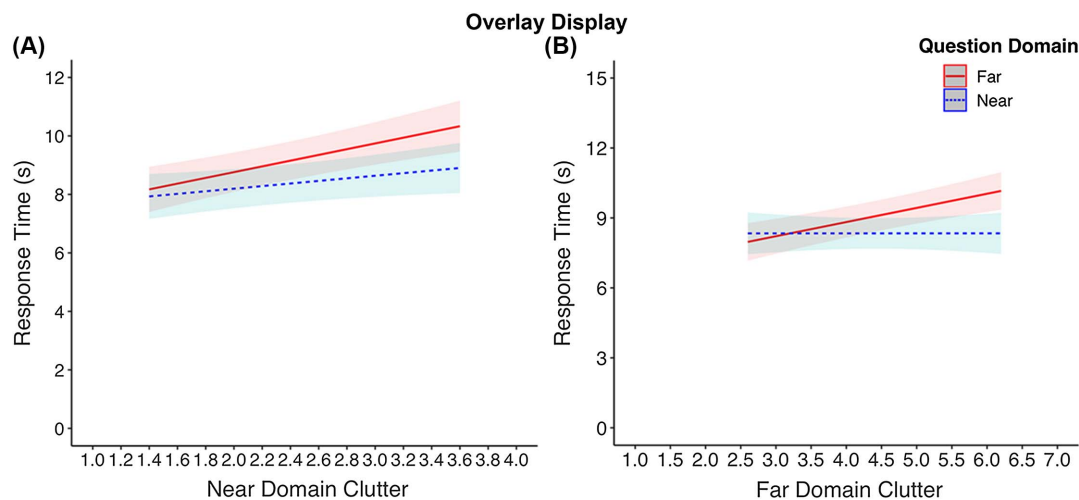
Overlay Display Response Time: Effects of Near- and Far-Domain Clutter

Near-Domain Clutter. Figure 4A presents the effects from the regression model of near-domain clutter, as quantified by the feature congestion metric, on response time for both near-domain (dashed blue line) and far-domain (solid red line) focused attention questions. Each regression line is based on the predicted values for all possible levels and values from the predictors in our model. Confirming Hypothesis 3A, response time increased as near-domain clutter increased for questions answered in both domains, estimate = 0.98 s, $SE = 0.23, t = 4.31, p < .001, \text{Cohen's } d = 0.22$. There was no effect of question type, estimate = 0.51 s, $SE = 0.77, t = 0.68, p = .50, \text{Cohen's } d = 0.11$. The interaction between near-domain clutter and question type was not significant, estimate = -0.54 s, $SE = 0.32, t = -1.70, p = .09, \text{Cohen's } d = 0.12$.

Far-Domain Clutter. The LMM was the same as before, except that far-domain clutter was used instead of near-domain clutter (Figure 4B). Confirming Hypothesis 3A, there was a significant effect of far-domain clutter on response time, estimate = 0.61 s, $SE = 0.13, t = 4.58, p < .001, \text{Cohen's } d = 0.43$, and a significant effect of question type on response time, estimate = 1.94 s, $SE = 0.98, t = 1.98, p = .048, \text{Cohen's } d = 0.13$. These effects are interpreted in the context of the significant interaction, estimate = -0.61 s, $SE = 0.22, t = -2.81, p = .005, \text{Cohen's } d = 0.13$. Increasing far-domain clutter slowed response time to far-domain questions (slope = 0.61, $SE =$

Figure 4

Experiment 1: Marginal Mean Response Time (Seconds) as a Function of Near (A) and Far (B) Domain Clutter for Focused Attention Question Domain as Predicted by the Model for the Overlay Display



Note. Domain clutter is based on the feature congestion metric. Shading represents 95% CIs as calculated from the model. CI = confidence interval. See the online article for the color version of this figure.

0.13, 95% CI [0.35, 0.87]), but not near-domain questions (slope = 0.0004, $SE = 0.17$, 95% CI [-0.33, 0.34]).

Collectively, the results support Hypothesis 3, showing that increasing both near- and far-domain clutter increased response time.

Adjacent Display Response Time: Effects of Near- and Far-Domain Clutter

In the adjacent display, the far-domain image was presented directly adjacent to the near-domain image (see Figure 1). As such, these data represent the absence of overlay clutter effects. Response time to questions relevant to the adjacent display was analyzed the same way as the overlay display.

Near-Domain Clutter. Figure 5 presents the effects of increasing near-domain clutter on response time for both question domains. The effect of near-domain clutter was not significant ($p = .46$). There was a significant effect of question type on response time, estimate = -3.93 s, $SE = 0.66$, $t = -5.98$, $p < .001$, Cohen’s $d = 1.01$. This effect is better interpreted in the context of the significant interaction, estimate = 1.52 s, $SE = 0.27$, $t = 5.57$, $p < .001$, Cohen’s $d = 0.39$. There was a cost to response time only for near-domain questions as near-domain clutter increased, slope = 1.38, $SE = 0.20$, 95% CI [0.99, 1.77]. Increasing near-domain clutter only impacted performance on near-domain questions. As shown in Figure 5, increasing clutter in the adjacent near domain did not impact performance when focusing attention on the spatially separated far domain to answer questions, slope = -0.14, $SE = 0.19$, 95% CI [-.51, 0.23].

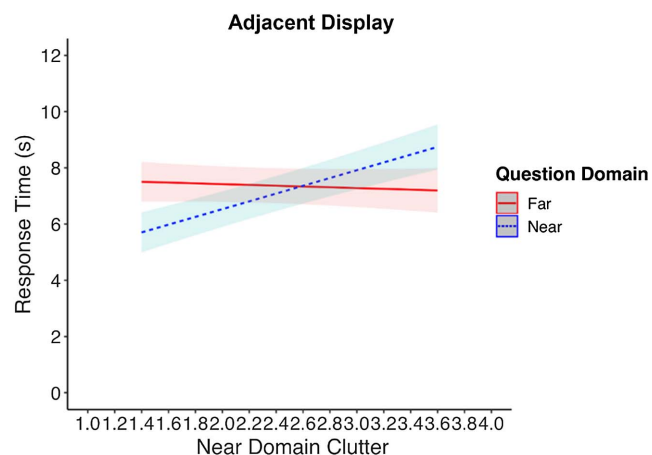
Far-Domain Clutter. When displays were adjacent, far-domain clutter had no significant effect on response time for questions in either domain, estimate = -0.02 s, $SE = 0.09$, $t = -0.16$, $p = .87$. There was also no significant effect of question type on response time (estimate = -1.22, $SE = 0.79$, $t = -1.55$,

$p = .12$) and no significant interaction (estimate = 0.17, $SE = 0.17$, $t = 1.05$, $p = .30$).

Collectively, when displays were overlaid, both near- and far-domain clutter significantly impacted response time for focused attention questions (Figure 4). However, when displays were adjacent, the only effect of clutter pertained to the near-domain clutter when answering near-domain questions. Therefore, the effects of near- and far-domain clutter were diminished when presenting the displays separately, which supports Hypothesis 3B regarding

Figure 5

Experiment 1: Marginal Mean Response Time (Seconds) as a Function of Near-Domain Clutter and Focused Attention Question Type Predicted From the Model for the Adjacent Display



Note. Near-domain clutter is based on the feature congestion metric. Shading represents 95% CIs as calculated from the model. CI = confidence interval. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

the interaction between the two sources of clutter (i.e., scene clutter and display clutter).

Effects of Feature Congestion Clutter on Integration Questions

For the integration task, data cannot be separated by near- and far-domain questions since the information from both domains was required to answer each question. We analyzed the data the same as described at the beginning of the Focused Attention Question section. However, display type (overlay vs. adjacent) was included as a fixed effect (submitted as a factor) in the models in replace of the type of focused attention questions.

Response Time

Near-Domain Clutter. Figure 6A depicts the influence of near-domain clutter on response time for the integration task. Participants took longer to answer integration questions for overlay compared to adjacent displays, estimate = 3.72 s, $SE = 0.75$, $t = 5.00$, $p < .001$, Cohen's $d = 0.59$. There was a significant effect of near-domain clutter on response time, estimate = 1.70 s, $SE = 0.22$, $t = 7.86$, $p < .001$, Cohen's $d = 0.27$. However, this effect is better explained by the significant interaction, estimate = -2.72 , $SE = 0.30$, $t = -8.92$, $p < .001$, Cohen's $d = 0.43$. When the displays were overlaid, increasing near-domain clutter caused a large increase in response time (slope = 1.69, $SE = 0.22$, 95% CI [1.27, 2.11]). When the displays were adjacent, increasing near-domain clutter decreased response time, slope = -1.02 , $SE = 0.25$, 95% CI [-1.45 , -0.60]. Supporting Hypothesis 3B, adjacent displays diminished the effects of increasing clutter compared to overlaid displays.

Far-Domain Clutter. Figure 6B depicts the effect of far-domain clutter on response time for the integration task. Increasing far-domain clutter increased response time, estimate = 0.46 s,

$SE = 0.14$, $t = 3.37$, $p < .001$, Cohen's $d = 0.07$. However, neither the effect of display type (estimate = -0.96 s, $SE = 0.90$, $t = -1.01$, $p = .29$, Cohen's $d = 0.15$) nor the interaction (estimate = -0.37 s, $SE = 0.20$, $t = -1.83$, $p = .07$, Cohen's $d = 0.06$) were significant.

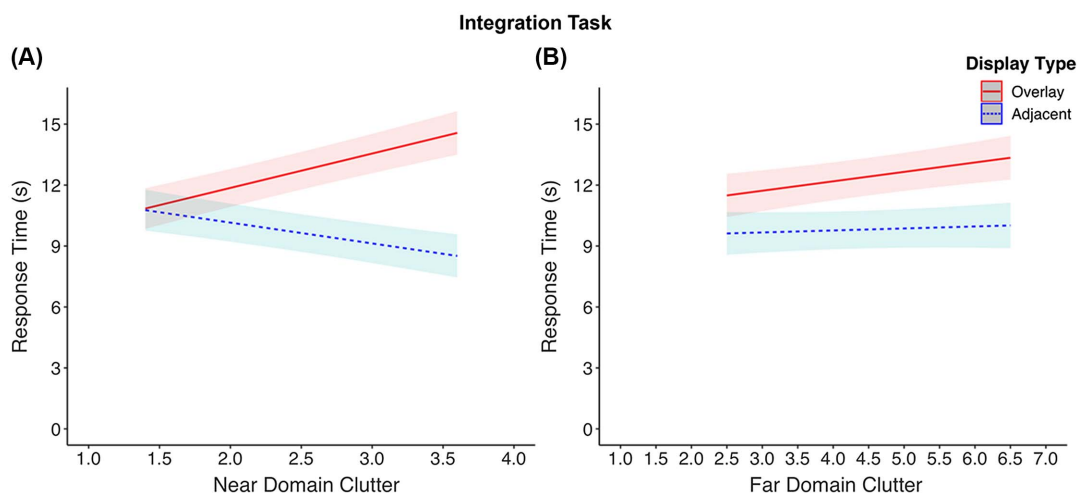
Error Rate Analyses

Error data were analyzed using binomial GLMM. In the GLMM, the dependent factor was error rate (coded as 1 for *incorrect* and 0 for *correct*), and the fixed effects were either near- or far-domain feature congestion clutter, question type (near- vs. far-domain focused attention questions), and their interaction. As with response time analyses, question type was submitted as a factor, and participant was included as a random effect with random intercepts. For the sake of brevity and because error data are generally less examined when investigating clutter effects, these data are only reported in Table 2. The main effect of clutter, question domain, and their interaction are reported for the focused attention task. The main effect of clutter, display type, and the interaction are reported for the integration task.

Overall, few effects of clutter on error rate were revealed for focused attention tasks. When the displays were overlaid, the results for error disconfirmed Hypothesis 3A, showing a decrease rather than an increase in error rate as clutter in either domain increased and show a speed-accuracy trade-off (SATO) that will be discussed in the General Discussion section. When the displays were adjacent, increasing near-domain clutter imposed a cost to both search time and accuracy for near-domain questions, suggesting an effect of computational numerosity clutter relevant to near-domain information.

For the integration task, error rate was higher in the adjacent condition (20%) compared to the overlay condition (15%). Also, the effects of clutter on error rate were dependent on the type of display. Notably, for overlay displays, neither near- nor far-domain clutter influenced error rate. However, when displays were adjacent,

Figure 6
Experiment 1: Marginal Mean Response Time (Seconds) as a Function of Near (A) and Far (B) Domain Clutter for Display Condition as Predicted by the Model for the Integration Task



Note. Near-domain clutter is based on the feature congestion metric. Shading represents 95% CIs as calculated from the model. CI = confidence interval. See the online article for the color version of this figure.

Table 2
The Binomial GLMM Results for the Focused Attention and Integration Task

| Focused attention | Clutter | Question | Interaction |
|-----------------------------|--------------------|-------------------|--------------------|
| Overlay display ND clutter | -2.80 ^a | -0.83 | 1.00 |
| Overlay display FD clutter | -5.38 ^a | -0.10 | 0.33 |
| Adjacent display ND clutter | -1.21 | -1.91 | 2.47 ^a |
| Adjacent display FD clutter | -1.27 | -1.23 | 1.68 |
| Integration | Clutter | Display | Interaction |
| ND clutter | 1.70 | 6.71 ^a | -5.88 ^a |
| FD clutter | -1.05 | 3.23 ^a | -2.46 ^a |

Note. ND and FD represent near domain and far domain, respectively. Main effects include clutter, question domain, and display type. For the focused attention task, negative signs for the main effects indicated that (a) increasing clutter reduced error rate, and (b) far-domain questions yielded fewer errors. For the integration task, negative signs for the main effects indicate that (a) increasing clutter reduced error rate, and (b) overlay displays yielded fewer errors overall. GLMM = generalized linear mixed models.

^aIndicates a statistically significant effect or interaction.

increasing clutter in either domain reduced the error rate. Considering both response time and error rate data for the integration task, these findings reveal an asymmetrical impact that increasing computational clutter has on performance: It either amplifies negative effects on response time for overlaid displays or induces a positive effect on error rate for adjacent displays.

Overlay Display: Effect of Near-Domain Color Similarity

To examine the effects of color similarity, we used two different versions of the near-domain display: Aircraft and flight path information were either colored red or green (see Figure 1). We

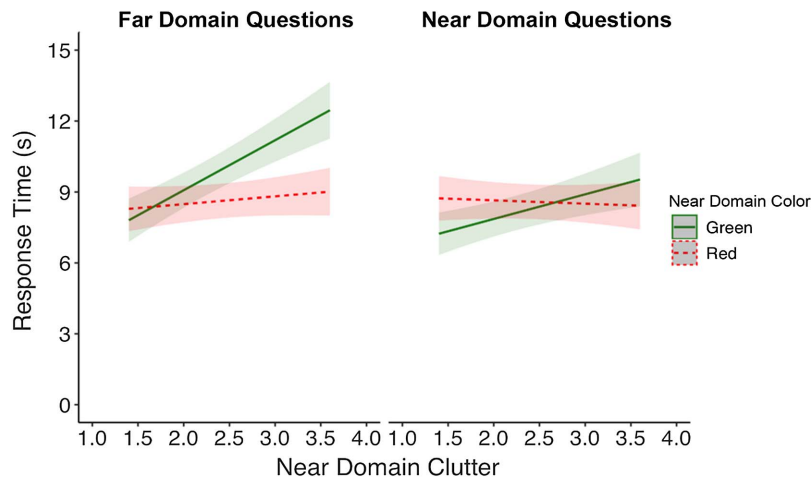
imposed this nested manipulation to examine any added clutter costs related to the color similarity between the two domains given that far-domain image scenes were generally greenish in hue (see Figure 1).

Hypothesis 4 predicted that greater color similarity (green) between the near-domain display and far-domain scene would impose greater performance costs to both focused attention tasks, specifically in the overlay display condition where computational clutter effects are prominent. We analyzed the data using the same LMMs described above but also including color (red vs. green) as a factor in the model. Figure 7 presents the effects of increasing display clutter on focused attention to the far domain (left panel) and the near domain (right panel) when colors between the domains were similar (green lines) and different (red lines), respectively.

There was a significant effect of color on response time, estimate = 2.98 s, SE = 1.11, *t* = 2.69, *p* = .007, Cohen’s *d* = 0.66. This effect is further explained by the significant interaction between color and near-domain clutter, estimate = -1.79 s, SE = 0.46, *t* = -3.88, *p* = .0001, Cohen’s *d* = 0.40. The slopes indicate a substantial increase in response time as green near-domain clutter increases (slope = 1.58, SE = 0.24, 95% CI [1.11, 2.04]), whereas response time remained relatively flat with increasing red near-domain clutter (slope = 0.09, SE = 0.21, 95% CI [-0.32, 0.51]). Confirming Hypothesis 4, performance costs on focused attention questions in either domain were greater when the near and far domains were similar in color.

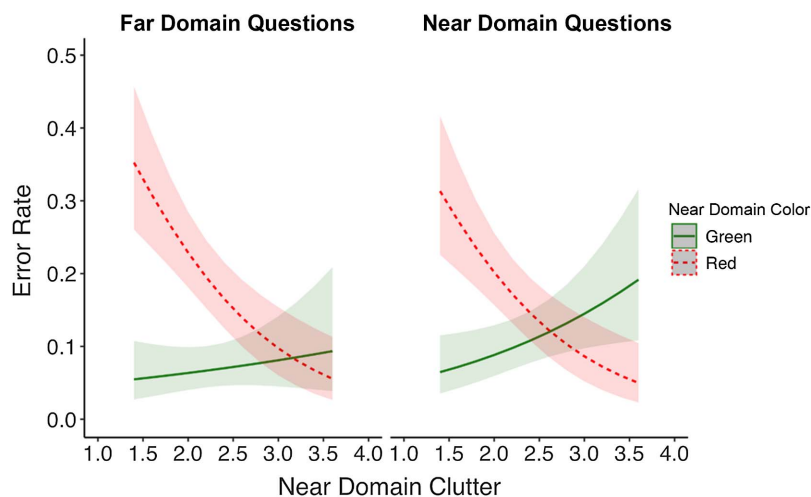
For error rate (Figure 8), we analyzed the data in the same way, except that we used a binomial GLMM. There was a significant effect of color on error rate, estimate = 4.02, SE = 0.95, *z* = 4.24, *p* < .001. Critically, there was a significant interaction between color and near-domain clutter, estimate = -1.27, SE = 0.54, *z* = -3.13, *p* = .002. For red displays (different color), as near-domain clutter increased, error rate decreased sharply, slope = -1.00, SE = 0.17, 95% CI [-0.13, -0.66]. Conversely, for green overlays

Figure 7
Experiment 1: Marginal Mean Response Time (Seconds) as a Function of Near-Domain Clutter and Color of Overlay (i.e., Near-Domain Image) Predicted From the Model for the Overlay Display



Note. Near-domain clutter is based on the feature congestion metric. Shading represents 95% CIs as calculated from the model. CI = confidence interval. See the online article for the color version of this figure.

Figure 8
Experiment 1: Marginal Mean Error Rate as a Function of Near-Domain Clutter and Color of Overlay (i.e., Near-Domain Image) Predicted From the Model for the Overlay Display



Note. Near-domain clutter is based on the feature congestion metric. Shading represents 95% CIs as calculated from the model. CI = confidence interval. See the online article for the color version of this figure.

(similar color), error rate slightly increased, slope = 0.41, SE = 0.21, 95% CI [0.01, 0.82]. As with response time, this effect was independent of whether the focused attention question was on the near or far domain.

Overall, increasing near-domain clutter with green icons imposes a greater performance cost for both response time and error rate irrespective of whether the questions are relevant to the near or far domain.

Clutter Metric Comparison

We compared the Rosenholtz feature congestion metric to the numerosity clutter metric applied to the near-domain display (i.e., the 2D electronic map) to assess whether one metric predicted performance better than the other. This was not done for the far domain because a value could not be computed for numerosity clutter. To compare these metrics, we examined response time on focused attention questions about the overlay condition only because this is where clutter effects appear to be most prominent. We repeated the LMM analysis described in the previous section for the feature congestion metric. The numerosity clutter metric used five quantitative levels of clutter based on the combination of low, medium, and high categories of both domains (see Tables 2 and 3). Additionally, we compared the object count (9, 16, and 29) numerosity clutter metric to the feature congestion metric.

For the three models described below, response time was the dependent factor. The fixed effects for the feature congestion metric included the feature congestion metric, question type (near or far domain focused attention questions), and their interaction. The fixed effects for the numerosity (“object count”) clutter metric included the rank-ordered clutter levels (Table 3) for one model and the “object count” for the other model, question type (near- or far-domain focused attention questions), and their interaction. Question type

was submitted as a factor in all models. Participant was included as a random effect in both models with random intercepts.

For the feature congestion metric, the effect of overall clutter was significant, estimate = 0.56 s, SE = 0.13, t = 4.30, p < .001, Cohen’s d = 0.12. In addition, there was a significant interaction between clutter and question type, estimate = -0.53 s, SE = 0.21, t = -2.49, p = .01, Cohen’s d = 0.12. Follow-up contrasts revealed that response time sharply increased for far-domain focused attention questions as overall clutter increased, slope = 0.56, SE = 0.13, 95% CI [0.31, 0.82], and remained relatively flat for near-domain questions, slope = 0.03, SE = 0.17, 95% CI [-0.30, 0.36]. This finding suggests that attention to far-domain information was more disrupted than to near-domain information.

For the numerosity clutter metric, the corresponding analysis revealed a significant effect of clutter, estimate = 1.12 s, SE = 0.14, t = 8.0, p < .001, Cohen’s d = 0.25. Critically, the results again showed a significant interaction between question type and the numerosity clutter metric, estimate = -0.90, SE = 0.20, t = -4.54, p < .001, Cohen’s d = 0.20. Follow-up contrasts revealed that response time increases sharply for far-domain questions, slope = 1.12, SE = 0.14, 95% CI [0.84, 1.39], but remained relatively flat

Table 3
Categorization of How Each Clutter Level Was Rank-Ordered Based on Near- and Far-Domain Clutter, Respectively

| Rank order | Clutter level |
|------------|--|
| 1 | (Low, low) |
| 2 | (Low, medium), (medium, low) |
| 3 | (Low, high), (high, low), (medium, medium) |
| 4 | (Medium, high), (high, medium) |
| 5 | (High, high) |

for near-domain questions, slope = 0.22, $SE = 0.14$, 95% CI [-0.05, 0.49].

Comparing the magnitude of the interaction suggests a greater effect with the numerosity clutter metric (Cohen's $d = 0.25$) than the Rosenholtz metric (Cohen's $d = 0.12$). We used the Bayesian information criterion (BIC) to assess which model better predicts performance. For the BIC, a difference of 2 is considered as positive evidence for one model over the other, and a difference greater than 10 suggests very strong evidence in favor of one model over the other (Fabozzi et al., 2014). The two models differed by 46.2, suggesting very strong evidence for the numerosity clutter metric model compared to the Rosenholtz feature congestion metric. We found similar Bayesian results when separately examining clutter in the near and far domains. For near- and far-domain clutter, the differences in BICs were 31 and 7, respectively, suggesting very strong and strong evidence, respectively, for the numerosity clutter model over the Rosenholtz feature congestion model. Based on these comparisons, the model using a combination of subjective categorization (far domain) and object count (near domain) had roughly an effect size twice as large as the feature congestion metric, for both the main effect (overall clutter effect) and interaction.

Next, we compared the object count (9, 16, and 29) of the near-domain numerosity clutter metric to the near-domain feature congestion clutter metric to see which model best predicted performance. For brevity, we only report the BICs. However, the analyses revealed a significant effect of near-domain clutter for both metrics ($p < .001$) but only a significant interaction between question type and clutter for the numerosity metric ($p = .02$). The model comparison revealed a difference in BICs of 17.8, again suggesting strong evidence for the object count numerosity clutter metric over the Rosenholtz feature congestion model. Taken together, these findings further highlight the effectiveness of the numerosity clutter metric.

Discussion

Experiment 1 addressed several hypotheses related to the quantitative and qualitative effects of clutter when examining overlaid displays, such as map overlays (Kroft & Wickens, 2002; Wickens & Ward, 2017). Our results focus more on response time, given that this is the dependent variable most often examined in search tasks (Moacdieh & Sarter, 2015; Rosenholtz et al., 2007).

Hypothesis 1 predicted longer response times with overlay displays compared to adjacent displays due to the dominant influence of overlay clutter on search time. This was confirmed for response time (Figure 2) and replicates effects in prior work (Kroft & Wickens, 2002). Figure 2 suggests that the reduced scanning in the overlay display was outweighed by the decrease in clutter of the adjacent display, resulting in faster response times for the adjacent display. We argue that the increased separation between the two domains for adjacent displays is sufficiently small, such that both domains remain within the eye field where information access costs (i.e., costs of scanning) are relatively small (Warden, Wickens, Clegg, Rehberg, et al., 2023). The overlay display, despite reduced scanning, exhibits higher LDC in each region when the domains are superimposed than when positioned side by side. This higher LDC remains in the overlay display but is eliminated in the adjacent display, which helps improve performance regardless of the increased visual scanning required. Notably, the scan-clutter trade-off shows comparable error rates for the two display conditions (Figure 3), suggesting that

adjacent displays lowered response time more than overlay displays without compromising accuracy.

Hypothesis 2 predicted that any costs or benefits associated with either display would be modulated by the type of attentional task (i.e., the proximity compatibility principle). Our results show a large cost of separation for both speed and accuracy for the more difficult integration task compared to the focused attention task. For accuracy, the interaction between display and task type for error rate strongly supported Hypothesis 2: overlay greatly improved integration accuracy but hindered focused attention accuracy (Figure 3). However, Hypothesis 2 was not supported for response time. Overlay substantially increased response time for both task types by a roughly equal margin (Figure 2). These two findings suggest a SATO in the data, an issue addressed in the General Discussion section.

Hypothesis 3 predicted that increasing clutter via both numerosity and feature congestion metrics would generally increase response time (Hypothesis 3A) and disproportionately so in the overlay display (Hypothesis 3B). Analysis of regression graphs (Figures 4 and 5) supported Hypothesis 3A: Across all analyses, the main effect of clutter on response time for focused attention questions was positive and significant for overlay displays. There was also strong support for Hypothesis 3B, with the average Cohen's d effect size indicating a much larger effect of clutter on performance for overlay displays ($d = 0.24$) compared to adjacent displays ($d = 0.06$), regardless of question type. For integration tasks alone, response time increased as far-domain clutter increased for both displays but decreased as near-domain clutter decreased for the adjacent display. Thus, while the computational clutter cost for integration questions was not amplified in overlay displays for the integration task, it was exclusive to that display.

Moreover, the amplifying effects of clutter are evident in the significant interaction between clutter domain (near or far) and focused attention question type. Averaged across display type, this interaction reveals that focused attention on the far domain is hurt more by increasing near-domain clutter ($d = 0.26$) than is focused attention on the near domain is hurt by increasing far-domain clutter ($d = 0.09$), an issue we address in the General Discussion section.

In Hypothesis 4, we also identified the important role of color similarity between near and far domains of the overlaid display in amplifying costs related to overlay clutter. Overlay displays using green for near-domain information incurred a greater performance cost as near-domain clutter increased compared to using red information for the near domain.

Finally, in Hypothesis 5, we compared the feature congestion clutter metric to the numerosity (object count) clutter metric. Overall, we found that the numerosity clutter metric predicted performance better than the feature congestion metric. The line and object count of our metric appears to have better captured clutter in the experiments than the decomposition of each domain into color, luminance, and orientation features achieved by the feature congestion metric. However, it is important to reemphasize that only the feature congestion metric can quantitatively capture the effects of scene clutter.

Experiment 2

Experiment 1 examined the scan-clutter trade-off but could not directly evaluate the effects of increased spatial separation on information integration or focused attention task performance due to the inherent confound of the adjacent condition, which combines

increased spatial separation with the removal of clutter. In Experiment 2, we address this by comparing performance in the adjacent condition with a condition where the two domains are separated by 32° of visual angle. This condition ensures sufficient spatial separation and evokes head movements (Warden et al., 2022). Neither of these conditions produces overlay clutter. Hypothesis 5, grounded in prior research (Draschkow et al., 2021; Poole et al., 2023; Schons & Wickens, 1993), predicts costs to performance for integration tasks due to the increase in IAE. Furthermore, the absence of overlay clutter from both adjacent and separated display conditions implies that the substantial clutter effects observed in Experiment 1 should be minimized at both of these display separations and only observed in an overlay condition, which is also included in Experiment 2.

Method

Participants

Fifty-one students from Colorado State University participated in the experiment using the same criteria for selection as those used in Experiment 1. Based on the effect size from Experiment 1, we increased the sample size to achieve sufficient power. No students from Experiment 1 participated in Experiment 2.

Stimuli and Apparatus

All stimuli were presented on a BenQ IPS monitor measuring 27 inc. on the diagonal with a screen resolution of $1,920 \times 1,020$ pixels and a 60 Hz refresh rate. All stimuli were created

the same, except that we added a third map condition, referred to as the separated condition, where the far- and near-domain images were separated by 32° (see Figure 9). To achieve this visual angle, participants were seated approximately 73 cm (28.75 inc.) away from the display.

Design

Participants completed six blocks. Blocks were counterbalanced based on task type (focused attention and integration) and display type (overlaid, adjacent, and separated). Each block consisted of 36 randomized experimental trials. Participants completed a total of 216 experimental trials plus 4 practice trials with auditory feedback for each display condition and task. Trial configurations were the same as in Experiment 1. The entire experiment lasted approximately 40 min.

Procedure

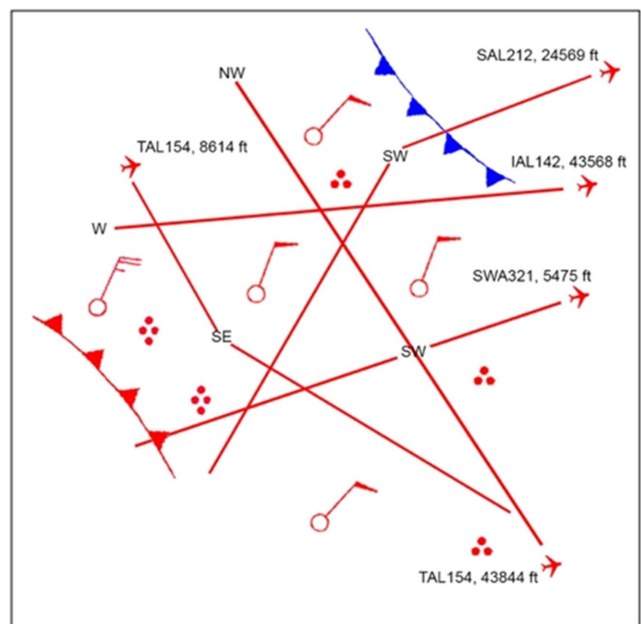
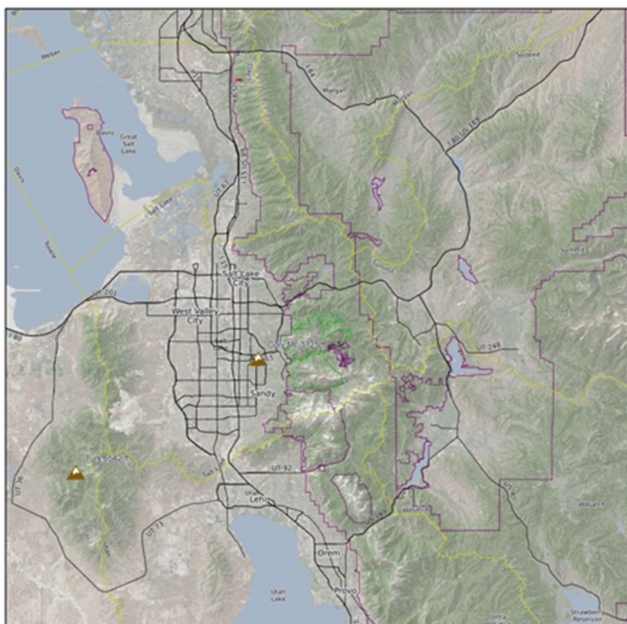
The procedure was the same as Experiment 1, except that they completed two additional blocks consisting of the separated display condition for each type of task (focused attention and integration).

Results

Before conducting the analyses, we looked for outliers using the same method described in Experiment 1. Only one participant had response time and accuracy data, suggesting they were an outlier and were excluded from the analysis. Five additional participants were excluded because they did not finish the entire experiment.

Figure 9

An Illustration of a Single Trial of the Separated Display Condition



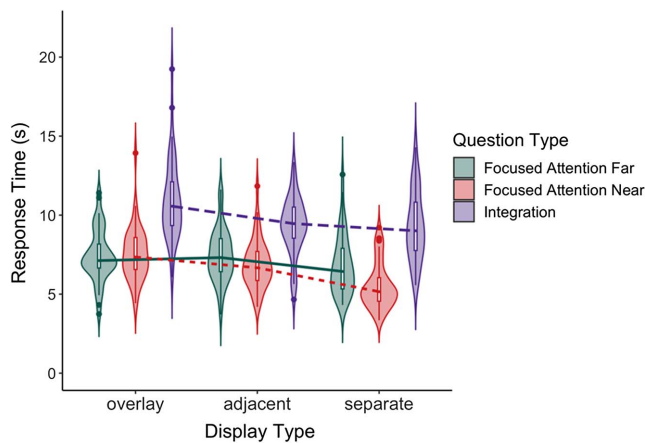
Note. The trial represents medium far-domain clutter (left) and high near-domain clutter (right). The far- and near-domain images were separated by 32° of visual angle, measured from the center of each image. Air route directions are abbreviated using the cardinal directions of North (N), South (S), East (E), and West (W). Aircraft identifiers (e.g. TAL, SWA) represent unique aircrafts. See the online article for the color version of this figure.

Effects of Display Separation: IAE Versus Overlay Clutter

Using two separate 2×3 repeated measures ANOVAs, we examined response time (Figure 10) and percent error (Figure 11) averaged across the different feature congestion clutter levels as a function of question type (focused near domain, focused far domain, and integration) and display type (overlay, adjacent, separate). For both ANOVAs, normality assumptions were checked the same way as in Experiment 1. Response time met assumptions of normality; however, percent error did not. Despite the removal of outliers and data transformations, this assumption was still violated. However, the ANOVA is robust to normality violations when sample sizes are moderate to large. Thus, the analysis reported here is still reliable given the moderate sample size ($N = 45$). In addition, the visual patterns and magnitude of effect sizes replicate findings from Experiment 1, suggesting reliability in the data. As in Experiment 1, performance in the overlay condition was predicted to be worse than in the adjacent conditions (Hypothesis 1). However, this effect was predicted to be modulated by the type of task (Hypothesis 2) and the spatial separation between the information displays (Hypothesis 6).

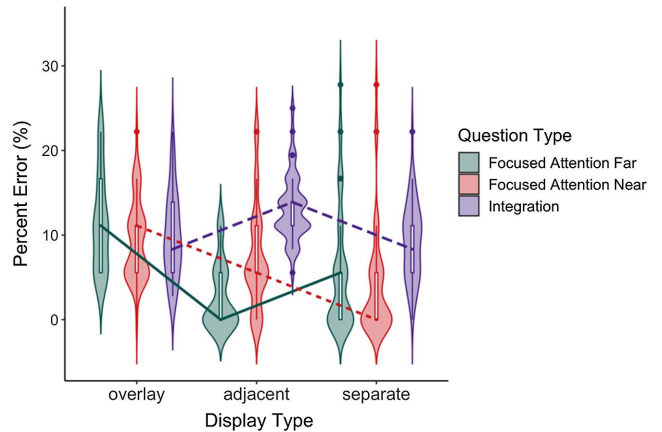
Response Time. The first ANOVA, shown by the left 6 points in Figure 10, examined the replication of Experiment 1 for the overlay and adjacent displays as a function of the task (Hypothesis 1 and Hypothesis 2). Reconfirming Hypothesis 1, results revealed a main effect of display type on response time, $F(1, 44) = 14.59, p = .0004, \eta_p^2 = 0.25$, indicating that response times were slower for the overlay condition compared to the adjacent condition. Participants were also slower with integration questions compared to focused attention questions, $F(2, 88) = 256.1, p < .001, \eta_p^2 = 0.85$. Additionally, there was a significant interaction between display type and question type, $F(2, 88) = 18.45, p < .001, \eta_p^2 = 0.30$. Overlaying displays slowed response time for integration questions more than for focused attention questions, $t(44) = -14.91, p < .001$, Cohen's $d = -1.62$. These findings replicate Experiment 1:

Figure 10
Experiment 2: Mean Response (Seconds) Time Plotted as a Function of Display Type and Question Type



Note. Green, red, and purple represent focused attention far domain, focused attention near-domain, and integration questions, respectively. Error bars represent one standard error of the mean. See the online article for the color version of this figure.

Figure 11
Experiment 2: Mean Percent Error Plotted as a Function of Display Condition and Question Type



Note. Green, red, and purple represent focused attention far-domain questions, focused attention near-domain, and integration questions, respectively. Error bars represent one standard error of the mean. See the online article for the color version of this figure.

Performance for integration questions was slower than that for focused attention questions, and the detrimental effect of overlay clutter was greater for integration than for focused attention. The results again contradict the response time predictions of Hypothesis 2.

The second ANOVA, represented by the right 6 data points in Figure 10, explored whether increasing IAE from adjacent to separated, unconfounded by clutter, increased response time in general and disproportionately so for the integration task (Hypothesis 6A). There was a significant effect of display separation on response time, $F(1, 44) = 10.33, p = .002, \eta_p^2 = 0.19$. Response time actually decreased as separation distance increased, disconfirming Hypothesis 6A. The effect of question type was significant, $F(2, 88) = 223.2, p < .001, \eta_p^2 = 0.84$. As seen before, response time was greater for integration than for focused attention questions. The interaction was significant, $F(2, 88) = 19.12, p < .001, \eta_p^2 = 0.30$, indicating that response time for focused attention questions was faster with greater separation, $t(44) = 5.14, p < .001$, Cohen's $d = 0.66$, but not different for integration questions, $t(44) = -0.06, p = .95$, Cohen's $d = -0.008$. Increasing separation seemed to impose no cost on integration questions and mildly benefited performance on focused attention questions. In particular, focusing attention on questions in the near domain benefited more from greater separation than did focusing attention on the far domain.

Error Rate. Error rate data were analyzed in the same way as response time data above and are shown in Figure 11. The ANOVA comparing overlay to adjacent display types (the 6 points on the left) revealed a significant main effect of display separation, $F(1, 44) = 23.09, p < .001, \eta_p^2 = 0.34$. Error rates were significantly higher with the integration than with the two focused attention tasks, $F(2, 88) = 24.15, p < .001, \eta_p^2 = 0.35$. However, these main effects are driven by the significant interaction, $F(2, 88) = 41.2, p < .001, \eta_p^2 = 0.48$. For overlaid displays, there were no differences in error rate based on the type of question, $t(44) = 0.52, p = .61$, Cohen's $d = 0.10$. However, in the adjacent condition, accuracy improved for focused attention questions but degraded for integration questions,

$t(44) = -11.28, p < .001$, Cohen's $d = -2.06$. These effects replicate those found in Experiment 1 and support Hypothesis 2.

The ANOVA comparing the adjacent to the separate display (the right 6 points) showed no significant effect of display separation, $F(1, 44) = 3.67, p = .06, \eta_p^2 = 0.08$. However, there was a significant effect of question type on error rate, $F(2, 88) = 62.1, p < .001, \eta_p^2 = 0.59$, where there were more errors on integration questions than on focused attention questions. The interaction between display type and question type was also significant, $F(2, 88) = 13.7, p < .001, \eta_p^2 = 0.24$. Increasing display separation improved performance on integration questions, $t(44) = 4.44, p < .001$, Cohen's $d = 0.92$, but did not improve performance on focused attention overall, $t(44) = 0.19, p = .85$, Cohen's $d = 0.04$. Increased separation improved accuracy when focusing attention on the near domain, $t(44) = 2.44, p = .02$, Cohen's $d = 0.51$, but reduced accuracy when focusing attention on the far domain, $t(44) = -2.59, p = .01$, Cohen's $d = 0.47$.

Overall, the findings revealed little evidence of a cost associated with increased separation, thereby contradicting the predictions of Hypothesis 6A. Interestingly, both the integration and focused

attention questions on the near domain showed a benefit of greater separation, particularly for response time to focused attention questions, and, for the integration task, for accuracy as well. It may be the case that presenting domains side by side (i.e., adjacent) introduces a source of visual distraction that greater separation seems to mitigate, an issue revisited in the following section.

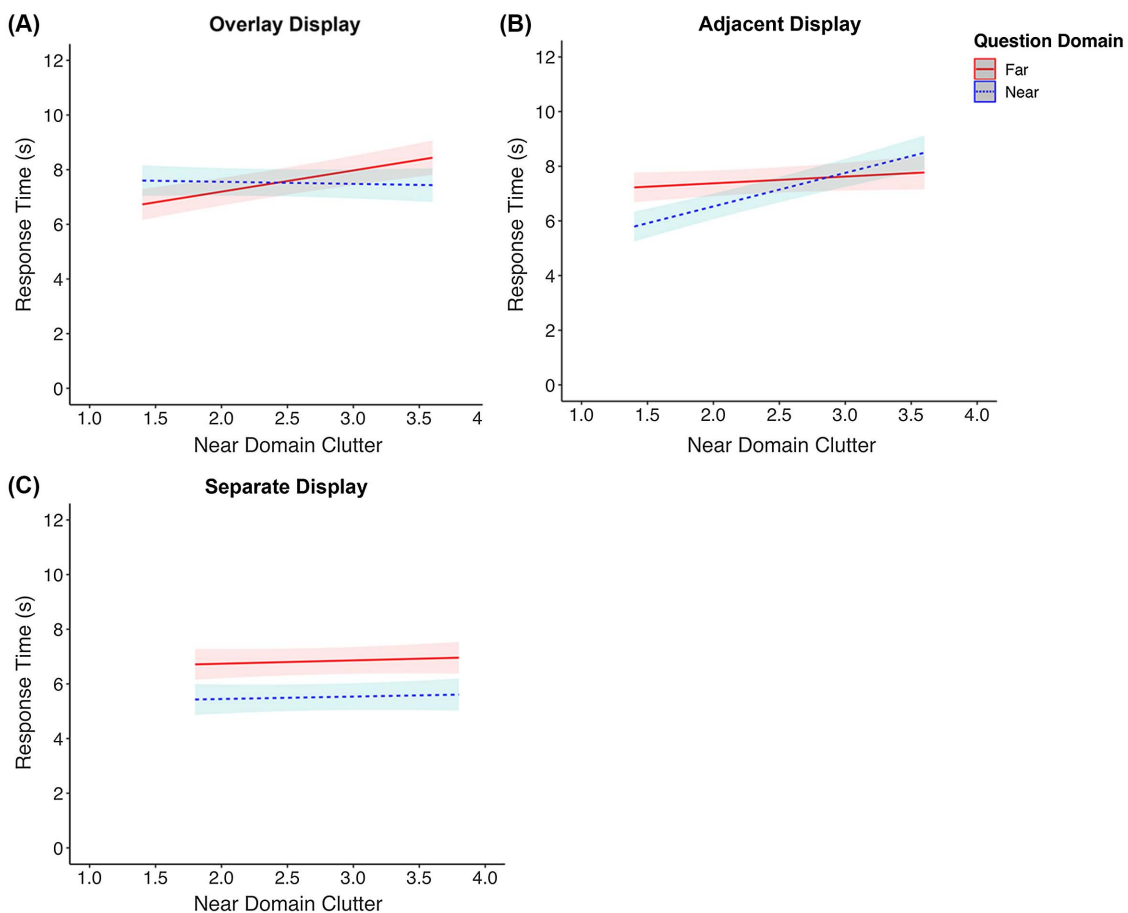
Effects of Feature Congestion Clutter

We analyzed the feature congestion clutter effect for the focused attention and integration task the same way as Experiment 1 (Figures 4–9), except that we only report the effects of near-domain clutter on response time because of the dominant clutter effects found in Experiment 1 (see Figures 4 and 5). Figure 12 depicts the most relevant data for comparing the effects of clutter for the overlay, adjacent, and separate displays.

In the overlay display condition (Figure 12A), there was a significant effect of near-domain clutter (estimate = 0.78 s, $SE = 0.15, t = 5.08, p < .001$, Cohen's $d = 0.22$) and question type (estimate = 2.07, $SE = 0.52, t = 4.00, p < .001$, Cohen's $d = 0.59$) on

Figure 12

Experiment 2: Marginal Mean Response Time (Seconds) as a Function of Near-Domain Clutter and Focused Attention Question Domain as Predicted From the Model for the Overlay (A), Adjacent (B), and Separate (C) Displays



Note. Near-domain clutter is based on the feature congestion metric. Shading represents 95% CIs as calculated from the model. CI = confidence interval. See the online article for the color version of this figure.

response time. However, these effects were driven by the significant interaction, estimate = -0.85 , $SE = 0.21$, $t = -4.01$, $p < .001$, Cohen's $d = 0.24$. Partially supporting Hypothesis 3, follow-up contrasts reveal that increasing near-domain clutter slowed response time to far-domain questions, slope = 0.78 , $SE = 0.15$, 95% CI [0.48, 1.08], while responses to near-domain questions were unaffected, slope = -0.08 , $SE = 0.15$, 95% CI $[-0.37, 0.21]$.

In the adjacent display condition (see Figure 12B), there was no significant effect of near-domain clutter on response time, estimate = 0.25 , $SE = 0.16$, $t = 1.52$, $p = .13$, Cohen's $d = 0.07$. However, near-domain questions were answered significantly faster than far-domain questions, estimate = -2.80 , $SE = 0.56$, $t = -4.97$, $p < .001$, Cohen's $d = 0.77$. The significant interaction between near-domain clutter and question type, estimate = 0.98 s, $SE = 0.23$, $t = 4.17$, $p < .001$, Cohen's $d = 0.27$, replicating the effects from Experiment 1 (Figure 5), signaled that increasing near-domain clutter slowed responses to near-domain questions (slope = 1.23 , $SE = 0.17$, 95% CI [0.88, 1.56]) more than far-domain questions (slope = 0.25 , $SE = 0.16$, 95% CI $[-0.07, 0.57]$).

In the separated display condition, the effect of question type was marginally significant, estimate = -1.23 , $SE = 0.62$, $t = -1.98$, $p = .05$, Cohen's $d = 0.35$, suggesting that far-domain questions took longer overall than near-domain questions, as seen in Figure 12C. Neither the effect of near-domain clutter ($p = .42$) nor the interaction ($p = .88$) was significant.

Finally, the effects of near-domain clutter and display separation on the integration task are shown in Figure 13. Results revealed a significant interaction between near-domain clutter and display condition such that increasing near-domain clutter imposes response time costs for overlay displays, estimate = 2.16 , $SE = .22$, $t = 10.04$, $p < .001$, Cohen's $d = 0.44$. However, response time is reduced by increasing clutter for both the adjacent (estimate = -2.16 , $SE = 0.22$, $t = -10.04$, $p < .001$, Cohen's $d = 0.44$) and separate (estimate = -2.42 , $SE = 0.21$, $t = -11.66$, $p < .001$, Cohen's $d = 0.50$) displays. This finding replicates the effect in Experiment 1 (Figure 6) and

disconfirms Hypothesis 5. Specifically, increasing near-domain clutter led to slower responses when displays were overlaid but faster responses when they were adjacent and separated.

We also completed the full set of clutter analyses that were performed in Experiment 1 on the Experiment 2 data. These are presented in tabular form representing the main effects of clutter, task type, and interaction in Appendix B.

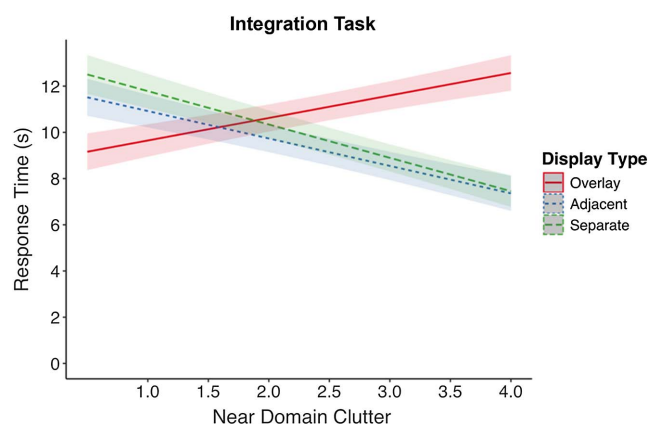
Color Analysis

Using the same analysis technique as Experiment 1, we examined the effect of near-domain color on response time for the overlay condition only as a function of the type of focused attention question. Hypothesis 4 predicted that greater color similarity would impose greater performance costs for both focused attention tasks. The effect of color significantly influenced response time, estimate = 3.88 , $SE = 0.73$, $t = 5.32$, $p < .001$, Cohen's $d = 1.34$. This effect is better explained by the significant interaction between color similarity and near-domain clutter, estimate = -2.05 , $SE = 0.30$, $t = -6.75$, $p < .001$, Cohen's $d = 0.60$. Replicating the results from Experiment 1, response time increases substantially as green (similar color) near-domain clutter increases, slope = 1.24 , $SE = 0.17$, 95% CI [0.93, 1.54], but not with the increase in red (different color) near-domain clutter. These findings, replicating those of Experiment 1, suggest that green overlay clutter is more disruptive than red, possibly due to the similarities with the terrain imagery.

Clutter Metric Comparison

In Experiment 2, we used the same BIC as in Experiment 1 to determine which model better predicts performance of the focused attention tasks in the overlay condition. Replicating the findings from Experiment 1, the two models differed by 25.2 for the overall clutter for each metric in the overlay condition, suggesting strong evidence for the numerosity clutter model over the Rosenholtz model in this context. The model comparisons for near-domain clutter of the feature congestion and numerosity rank order clutter metric showed BICs that differed by 13.5, suggesting strong evidence in favor of the numerosity clutter metric. Last, the model comparison between the near-domain feature congestion and numerosity object count clutter metric showed BICs that differed by 3.55, suggesting positive evidence in favor of the numerosity clutter metric.

Figure 13
Experiment 2: Marginal Mean Response Time (Seconds) as a Function of Near-Domain Clutter and Each Display Condition as Predicted From the Model



Note. Near-domain clutter is based on the feature congestion metric. Shading represents 95% CIs as calculated from the model. CI = confidence interval. See the online article for the color version of this figure.

Discussion

Experiment 2 accomplished two fundamental goals. First, the results closely replicated key findings from Experiment 1 regarding the overall joint effects of overlay clutter and IAE, concerning the scan-clutter trade-off (Figures 10 and 11) between the adjacent and overlay displays. In addition, these results replicated the quantitative computational effects of near- and far-domain clutter (assessed by the feature congestion metric) in the overlay and adjacent displays as these interacted with the qualitative feature of overlay clutter on the three tasks (Figure 12).

The second goal was to examine the role of IAE, unconfounded with overlay clutter, by contrasting the performance of the same tasks between an adjacent and more separated display layout (Hypothesis 6). We observed the surprising absence of large costs of the more distant layout and also revealed the unpredicted negative influence

of clutter in adjacent displays (Figure 12, in essence disconfirming Hypothesis 6). Furthermore, in some cases, we found that increasing display separation actually reduced response time as clutter increased, suggesting some benefits of increased display separation to produce some space away from the unwanted clutter.

General Discussion

Complex visual information, such as maps, can be presented with multiple databases of information that are either directly overlaid onto the same visual space or further separated (e.g., presented side by side or above and below). How these different databases are presented can impact performance when searching for information relevant to either a single database or searching and comparing information in both, due to the trade-off in costs associated with increasing clutter in one or both databases and scanning between sources of information. In two experiments, we explored the role of clutter in overlaid displays, typical of many digital map and weather programs, such as onX Maps (onX, 2023), CalTopo (CalTopo, 2021), the National Digital Forecast Database (National Centers for Environmental Information, 2023), and those used in HUDs and HMDs. We also examined how this clutter might trade off against IAE when the same two displays or sources of visual information are presented separately (i.e., either adjacent or at a more separated visual angle). This trade-off is referred to as the scan-clutter trade-off and further proposes that when information between two sources has to be integrated, information should be presented closer together. However, when only one source of information is required, then that information can be presented further apart, increasing the IAE. We assumed that this trade-off would be moderated by the amount of clutter quantified within each display and by the type of attentional task. Below, we address six hypotheses regarding this trade-off.

Hypothesis 1, proposing that overlay clutter would increase response time, was confirmed in both experiments (Figures 2 and 12) and represents a direct extrapolation of basic visual search research. That is, as the set size increases, search time increases linearly (e.g., Wolfe, 2021). For all tasks in the current work, at least one component of performance is always impacted by searching for the element(s) prompted to be searched for in the question, and this search is harmed by overlay clutter. Further confirmation of the hypothesis is provided by the consistent effect of increasing computational clutter on increasing response time in all of the overlay displays (Figures 4, 6, and 12). Regarding the scan-clutter trade-off, the response time results confirm that costs associated with clutter play a greater role than costs associated with increased IAE. We speculate, and as observed in other research (Warden, Wickens, Clegg, Rehberg, et al., 2023), that the cost of eye movements underlying IAE in the adjacent condition is actually quite low in terms of effort and, therefore, does not impose performance costs. Our results are also consistent with the findings concerning HUDs (Fadden et al., 2000) and HMDs (Warden, Wickens, Clegg, & Ortega, 2023), showing that when focusing attention on one domain or the other to detect low-resolution targets was required, overlaying information was harmful.

Hypothesis 2 tested the proximity compatibility principle's (Wickens & Carswell, 1995) prediction that any costs associated with presenting information on adjacent displays will be exaggerated to a greater extent for a focused attention task relative to an integration task. Prior work examining database overlay (Kroft & Wickens, 2002; Wickens & Ward, 2017) suggests the harmful effects of

overlay clutter should be mitigated or offset to the extent that information integration is required between the overlaid sources of information because the close spatial proximity of overlaying information minimizes any IAE costs for tasks that impose greater demands on working memory (e.g., "Is a certain aircraft north of a mountain?"). Surprisingly, for response time, the interaction between the type of task and the display was disconfirmed. In both experiments, the time cost of overlay was slightly but significantly amplified, not reduced, for the integration task compared to the focused attention tasks (Figures 2 and 10). While this effect was not predicted, it may be explained by the double search requirements of the integration task, which we describe below in the context of the SATO. In contrast, for accuracy, the interaction between the type of task and display predicted in Hypothesis 2 was confirmed in both experiments (Figures 3 and 11): There was an advantage of separated displays for focused attention questions but a cost when integration between the two domains was required. That is, overlaid displays supported performance more for integration questions than did separate displays. We may conclude that predictions of the proximity compatibility principle pertaining to database overlay are not universal and that there may be SATOs involved depending on the nature of the task and the displays. These findings suggest that designers should consider whether speed or accuracy is more important in the displayed tasks involved. We address the SATO in our results in more detail below.

Hypothesis 3 examined the specific effects of the computational clutter metrics, as quantified both by the feature congestion metric, which could be applied to both scene and display (Rosenholtz et al., 2007) and by the simpler object count metric of numerosity clutter, which is only applicable to human-rendered displays (Yeh & Wickens, 2001). In Hypothesis 3A, we hypothesized that increasing clutter by either metric would degrade performance, a hypothesis that was inconsistently confirmed across all of the regression results (Figures 4–6, and Figures 12 and 13). There was not a consistent cost to response time (i.e., a positive slope in the figures) and sometimes accuracy showed a benefit (i.e., a negative slope). However, the qualification offered by Hypothesis 3B accounted for this inconsistency by proposing that the detrimental effects of increasing computational clutter are primarily restricted to overlay displays. That is, there is an interaction between the quantitative (metric) aspects of clutter, associated with numerosity and/or feature congestion, and the qualitative aspect of clutter, associated with overlay. A similar interaction was observed by M. R. Beck et al. (2012), and a correspondence can be drawn between their manipulations of both "global density clutter" (C3) and "local density clutter" (i.e., proximity of clutter to the target of interest) on the one hand and our manipulations of computational clutter and overlay clutter, respectively, on the other. Prior work found that LDC had a greater influence on performance when GDC was high compared to low (M. R. Beck et al., 2012). In the context here, overlaying displays increases the LDC as the near and far domains are superimposed, causing a greater impact on performance than increasing computational clutter (GDC) alone. In short, overlay will amplify the negative effects of increasing computational clutter.

Hypothesis 4 addressed the specific degrading role of color similarity between the dominant colors of the two domains, reflecting the well-established performance cost of high color similarity between target and distractors in visual search (Geisler & Chou, 1995; Nagy & Sanchez, 1990; Wickens et al., 2022; Wolfe, 2021).

This was confirmed in both experiments. Establishing this degrading role of color similarity is important in the application of clutter modeling because the generally useful feature congestion model, while accounting for the color diversity of the images, does not account for and, in fact, predicts the opposite effect of color similarity. Our current results regarding the effect of color differences (Figures 7 and 8) showed a detrimental effect of color similarity on response time and a benefit to accuracy with color dissimilarity as computational clutter increased. Such results speak to the need to assess the dominant wavelength of information that may be overlaid via HUD or HMD applications and, for design applications, to consider the possible advantages of adapting display image hue contingent upon the dominant hue in the far-domain scene imagery (e.g., a green forest, a blue sky, or an orange desert).

Hypothesis 5 predicted the validity of both computational clutter metrics (feature congestion and numerosity) in predicting task performance (particularly response time). Our results revealed that both metrics predicted the time costs to performance and, in the conditions in which both metrics could be applied (near-domain display clutter), the object count metric of numerosity clutter outperformed the feature congestion metric by a substantial margin using a Bayesian analysis. The superiority of the numerosity clutter metric in predicting performance in near-domain clutter conditions suggests that different clutter metrics might be more relevant for specific types of tasks or information displays. For reasons described above, other metrics were not examined in the current research.

In Experiment 2, we addressed Hypothesis 6, which focused on the role of spatial separation or IAE in processing multiple visual displays. Hypothesis 6 predicted a cost to performance as spatial separation increased from 10° to 32°. However, we found that the performance cost due to increased spatial separation was muted, even in the integration task, which imposed a higher working memory load than the focused attention task. Two factors may explain this outcome. First, as has been shown in recent research, the negative performance consequences of spatial separation may have been overestimated in previous research because of highly effective compensatory head movements (K. H. Kim et al., 2010; Poole et al., 2023; Warden et al., 2022). In addition, eye and head movements are low-effort movements that can be conducted over smaller visual angles (less than 50°) to perceptually access information (Warden et al., 2022) compared to computationally expensive memory strategies that may degrade performance at greater visual angles (Ballard et al., 1995).

Second, and somewhat unexpectedly, we found that performance in the spatially “close” display condition (i.e., the adjacent condition) was, in fact, penalized by increasing computational clutter of the adjacent display. This effect, like the “flanker effect” observed in basic visual attention research (Eriksen & Eriksen, 1974), a hallmark of space-based theories of visual attention (LaBerge & Brown, 1989; Wickens et al., 2022), would have penalized performance in the adjacent condition but not in the more separated condition, thereby attenuating any performance cost in the separate relative to the adjacent condition. This effect can be seen in Figure 5 and Table 2, where increases in near-domain clutter disrupt both the speed and accuracy of questions about the adjacent far-domain information. Specifically, increasing clutter in the adjacent condition induced distracting elements when focusing on only one domain was required, interfering with task performance, whereas, in the separate condition, the two domains were sufficiently far apart to

eliminate the distracting elements of the irrelevant domain, thereby reducing response time. Examination of Figure 12 suggests that computational clutter effects were, in fact, impacted by increased spatial separation, such that the negative effects of adjacent computational clutter were eliminated in the more separated condition. In the context of designing information displays, careful consideration should be given to the relationship between spatial separation and computational clutter.

The Speed–Accuracy Trade-Off

One phenomenon, not incorporated in our hypotheses, was the unexpected appearance of SATOs (Drury, 1994; Pachella, 1974) in two aspects of our data. Here we had, perhaps naively, assumed that influences that would slow processes would also damage their accuracy. However, this was not the case.

The first SATO was expressed in the fundamental scan-clutter trade-off seen in Figures 2 and 3 and again in Figures 10 and 11. Figure 2 reveals that the clutter created by overlay displays slowed integration task performance slightly more (about ½ s more) than it slowed the performance of the focused attention task. We can explain, in hindsight, that this slowing is due to the integration task requiring visual search for two elements (one in each domain), whereas focused attention task requires search for only one element. The integration task effectively doubles the number of searches in a cluttered field, which increases search time. The SATO was then reflected in the opposite effect on accuracy (Figure 3), where overlay eliminated an accuracy cost on the memory-dependent integration task, but there was no memory cost of overlay to be eliminated on the “memory-lite” focused attention task. It is quite plausible to assume that memory effects specific to integration will be expressed in accuracy measures. Thus, this explanation of the trade-off effects in Figures 2 and 3 and also in Figures 10 and 11 is based on different influences on the two different processes involved in the scan-clutter trade-off.

The second appearance of the SATO is inferred to be more strategic than process-oriented. The expression of the SATO here is in the effects of computational clutter, whose increases consistently harmed speed more than accuracy and sometimes actually improved accuracy (Table 2), but rarely ever improved speed. To account for these results, we assume that, in the face of higher clutter, which is a visually salient attribute of an image, participants choose to slow down in order to preserve accuracy, and sometimes this strategy actually overcompensates to improve accuracy. That is, the SATO here is driven by a strategy choice.

Near–Far Asymmetry

Throughout the experimental results, we have observed a sort of asymmetry whereby near-domain clutter tends to hurt the focus of attention on the far domain, more so than the reverse. This asymmetry may be seen in Table A1 (Appendix A). For information integration questions, the effect sizes for manipulating near-domain display clutter are much greater for both overlay ($d = 0.25$) and adjacent ($d = 0.18$) displays compared to manipulating far-domain scene clutter for overlay ($d = 0.07$) and adjacent ($d = 0.02$) displays. Also, when comparing the graphs for the overlay display, we see, in Figure 4A, that near-domain clutter has a greater detrimental effect on far- than near-domain questions. However, in Figure 4B,

when far-domain clutter is increased, this pattern does not reverse. Again, just as with near-domain clutter, far-domain questions suffer more than near-domain questions. We might refer to this as a general “far domain vulnerability” to overlay, where the visual system prioritizes information that is perceptually closer in space. We note that this is not a result of the simple occlusion of a “closer” image over a “further” image (the images were carefully crafted to avoid such occlusion). Instead, it may reflect the same domain biases reviewed by [Bufacchi and Iannetti \(2018\)](#), in which greater urgency is placed, for evolutionary reasons, on events, objects, and movements substantially closer to the observer. For example, a snake within striking distance is more relevant than one further away. It would seem reasonable to infer that this bias, in the present context, could be reflected in the greater salience (and hence disruption) of clutter that is perceived as closer to the observer than the far-domain scene lying behind and thus farther away. This effect has design implications for HUDs and HMDs, which present information that overlays the far-domain scene directly in front of the user.

Practical Implications

The present work offers practical implications concerning design guidelines for multifaceted information displays such as digital maps that utilize multiple databases of information, HUDs, and HMDs. Based on the findings from both experiments, the performance cost of clutter had a greater impact than any costs associated with increasing display separation between sources of information, specifically when tasks required focused attention on one domain. This finding suggests that when speed is more important than accuracy, placing information on separate displays is a viable option for minimizing the effects of overlay clutter. In the context of 2D flat panel displays, this suggests that presenting map information on different monitors may be beneficial in scenarios when the amount of clutter is extensive and a decision must be made quickly (e.g., weather forecasting), particularly when information integration is not required. In the context of HMDs, these findings suggest that presenting information at the edge or even outside of the field of view of the device, as in the case of “Glanceable AR” (see [Lu et al., 2020](#)), may also be another viable option to avoid costs of overlay clutter. For tasks that require quick decisions in multidisplay environments, such as those used in command centers, control rooms, or the cockpit of a plane, the relationship between spatial separation and computational clutter should be considered. However, when the accuracy of information integration is required, then overlay becomes preferable.

Our findings also suggest that when information from more than one source has to be overlaid (e.g., display information presented with an augmented-reality HMD), specific colors may help performance on tasks compared to other colors. In the context here, using green, a color similar to the far-domain scene image, for the near-domain information hindered performance more than using red, a color dissimilar to the far-domain scene image. This finding has practical applications for the design of information displays, especially those that directly overlay information, such as HUDs or HMDs. Such displays present information on a far-domain scene that changes color as the user moves through the world. Therefore, designers should consider adapting the hue of the displayed information based on the dominant, and fairly enduring, hue in the far-domain scene imagery. In terms of design principles, this

finding suggests that caution and care should be taken when deciding which color to use to display information.

Last, our findings show the validity of computational clutter metrics in predicting task performance. Designers can leverage these clutter metrics to optimize information displays that consist of multiple sources of information. By using metrics like the object count of numerosity clutter, designers can minimize cognitive load and therefore enhance task performance.

Overall, we found reliable and replicable patterns in our results that provide us with insights into the effects of overlay clutter and IAE when reading displays with more than one layer of information. Understanding how to mitigate costs associated with overlay clutter is a critical factor in terms of design guidelines for HUDs, HMDs, and even 2D flat panel displays that present multiple databases overlaid together.

Limitations and Future Directions

The current work is not without limitations. Due to time constraints, the current work did not include the correlated feature of edge density clutter metric as a measure. Given the nature of our near-domain imagery, this measure may predict performance similar to or better than our count clutter metric. Future work should seek to incorporate this measure of clutter to further compare different clutter metrics. Another limitation is that we only tested two different colors for the near-domain image. Future work will seek to understand not only additional colors but also the role that luminance and contrast play in overlay clutter ([Ververs & Wickens, 1998](#)), especially in the context of HMDs and dynamic situations. We also did not collect eye-movement or head-movement data which could have given us more empirical insights about the role of IAE in terms of the amount of eye scanning and head movements. In order to ascertain the generalizability of the proximity compatibility principle, future work should also examine different types of integration tasks and tasks that impose greater working memory demands. In addition, the near-far domain asymmetry resulting in the “far domain vulnerability” warrants further investigation in order to understand the observed effect and identify ways to mitigate the negative effect of clutter on far-domain tasks. Last, to further improve the external validity of our findings, future work must examine the effects in the context of different types of visual search tasks (medical diagnosis or military operations), information displays (HMDs or HUDs), visual information (medical images or financial data), and spatial layouts (multiple displays).

Conclusions

Overall, we found reliable and replicable patterns in our results that provide us with insights into the effects of overlay clutter and IAE for displays with multiple sources of information. Design decisions to overlay information or present information separately depend on the tradeoff between IAE and clutter. Our findings highlight the importance of minimizing costs associated with overlay clutter to improve performance. Our results also reveal that, with the assistance of quantifiable clutter models such as feature congestion metric and object count metric of numerosity clutter, it is possible to design near-domain displays with clutter levels at which performance with overlay displays is equivalent to adjacent displays. This is particularly the case for tasks involving focused attention on the far

domain, which is directly related to the use of HUDs and HMDs that present information that overlays the real world. This latter conclusion is particularly important because it is here, in detecting targets through an overlaid display, that the HUD vulnerabilities to clutter have emerged (Fadden et al., 2000, 2001). Additionally, designers should carefully select the appropriate colors of overlaid information to enhance performance. Mitigating the costs of clutter is crucial for effective design, from HUDs and HMDs to 2D flat panel displays with overlaid databases.

References

- Andersen, E., Goucher-Lambert, K., Cagan, J., & Maier, A. (2021). Attention affordances: Applying attention theory to the design of complex visual interfaces. *Journal of Experimental Psychology: Applied*, 27(2), 338–351. <https://doi.org/10.1037/xap0000349>
- Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, 7(1), 66–80. <https://doi.org/10.1162/jocn.1995.7.1.66>
- Beck, M. R., Lohrenz, M. C., & Trafton, J. G. (2010). Measuring search efficiency in complex visual search tasks: Global and local clutter. *Journal of Experimental Psychology: Applied*, 16(3), 238–250. <https://doi.org/10.1037/a0019633>
- Beck, M. R., Trenchard, M., Van Lamsweerde, A., Goldstein, R. R., & Lohrenz, M. (2012). Searching in clutter: Visual attention strategies of expert pilots. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 56(1), 1411–1415. <https://doi.org/10.1177/1071181312561400>
- Broadbent, D. E. (1982). Task combination and selective intake of information. *Acta Psychologica*, 50(3), 253–290. [https://doi.org/10.1016/0001-6918\(82\)90043-9](https://doi.org/10.1016/0001-6918(82)90043-9)
- Bufacchi, R. J., & Iannetti, G. D. (2018). An action field theory of peripersonal space. *Trends in Cognitive Sciences*, 22(12), 1076–1090. <https://doi.org/10.1016/j.tics.2018.09.004>
- Bundesden, C., & Pedersen, L. F. (1983). Color segregation and visual search. *Perception & Psychophysics*, 33(5), 487–493. <https://doi.org/10.3758/BF03202901>
- CalTopo. (2021). *Backcountry mapping evolved*. <https://caltopo.com/>
- Dey, A., Billinghurst, M., Lindeman, R. W., & Swan, J. E., II. (2018). A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. *Frontiers in Robotics and AI*, 5, Article 37. <https://doi.org/10.3389/frobt.2018.00037>
- Donderi, D. C. (2006). Visual complexity: A review. *Psychological Bulletin*, 132(1), 73–97. <https://doi.org/10.1037/0033-2909.132.1.73>
- Draschkow, D., Kallmayer, M., & Nobre, A. C. (2021). When natural behavior engages working memory. *Current Biology*, 31(4), 869–874.e5. <https://doi.org/10.1016/j.cub.2020.11.013>
- Drury, C. G. (1994). The speed—Accuracy trade-off in industry. *Ergonomics*, 37(4), 747–763. <https://doi.org/10.1080/00140139408963686>
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3), 433–458. <https://doi.org/10.1037/0033-295X.96.3.433>
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16(1), 143–149. <https://doi.org/10.3758/BF03203267>
- Fabozzi, F. J., Focardi, S. M., Rachev, S. T., & Arshanapalli, B. G. (2014). *The basics of financial econometrics: Tools, concepts, and asset management applications* (pp. 399–403). Wiley. <https://doi.org/10.1002/9781118856406>
- 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 attention perspective and a meta analysis*. SAE Technical Papers Series. SAE International. <https://doi.org/10.4271/2000-01-5542>
- Geisler, W. S., & Chou, K. L. (1995). Separation of low-level and high-level factors in complex tasks: Visual search. *Psychological Review*, 102(2), 356–378. <https://doi.org/10.1037/0033-295X.102.2.356>
- Gray, W. D., & Fu, W. T. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science*, 28(3), 359–382. <https://doi.org/10.1016/j.cogsci.2003.12.001>
- Henderson, J. M., Chanceaux, M., & Smith, T. J. (2009). The influence of clutter on real-world scene search: Evidence from search efficiency and eye movements. *Journal of Vision*, 9(1), Article 32. <https://doi.org/10.1167/9.1.32>
- Holahan, C. J., Culler, R. E., & Wilcox, B. L. (1978). Effects of visual distraction on reaction time in a simulated traffic environment. *Human Factors*, 20(4), 409–413. <https://doi.org/10.1177/001872087802000404>
- Kaber, D. B., Alexander, A. L., Stelzer, E. M., Kim, S. H., Kaufmann, K., & Hsiang, S. (2008). Perceived clutter in advanced cockpit displays: Measurement and modeling with experienced pilots. *Aviation, Space, and Environmental Medicine*, 79(11), 1007–1018. <https://doi.org/10.3357/ASEM.2319.2008>
- Kim, K. H., Reed, M. P., & Martin, B. J. (2010). A model of head movement contribution for gaze transitions. *Ergonomics*, 53(4), 447–457. <https://doi.org/10.1080/00140130903483713>
- Kim, S. H., Prinzel, L. J., Kaber, D. B., Alexander, A. L., Stelzer, E. M., Kaufmann, K., & Veil, T. (2011). Multidimensional measure of display clutter and pilot performance for advanced head-up display. *Aviation, Space, and Environmental Medicine*, 82(11), 1013–1022. <https://doi.org/10.3357/ASEM.3017.2011>
- Kroft, P., & Wickens, C. D. (2002). Displaying multi-domain graphical databases: An evaluation of scanning clutter, display size, and user activity. *Information Design Journal*, 11(1), 44–52. <https://doi.org/10.1075/idj.11.1.06kro>
- LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review*, 96(1), 101–124. <https://doi.org/10.1037/0033-295X.96.1.101>
- Lazaro, M. J., Kang, Y., Yun, M. H., & Kim, S. (2021). The effects of visual complexity and decluttering methods on visual search and target detection in cockpit displays. *International Journal of Human-Computer Interaction*, 37(7), 588–600. <https://doi.org/10.1080/10447318.2021.1890491>
- Lee, J., Lee, N., Ju, J., Chae, J., Park, J., Sung Ryu, H., & Seok Cho, Y. (in press). Visual complexity of head-up display in automobiles modules attentional tunneling. *Human Factors*.
- Lohrenz, M. C., Trafton, J. G., Beck, R. M., & Gendron, M. L. (2009). A model of clutter for complex, multivariate geospatial displays. *Human Factors*, 51(1), 90–101. <https://doi.org/10.1177/0018720809333518>
- Lu, F., Davari, S., Lisle, L., Li, Y., & Bowman, D. A. (2020). *Glanceable AR: Evaluating information access methods for head-worn augmented reality* [Conference session]. 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Atlanta, GA, United States. <https://doi.org/10.1109/VR46266.2020.00113>
- Moacdieh, N., & Sarter, N. (2015). Display clutter: A review of definitions and measurement techniques. *Human Factors*, 57(1), 61–100. <https://doi.org/10.1177/0018720814541145>
- Nagy, A. L., & Sanchez, R. R. (1990). Critical color differences determined with a visual search task. *Journal of the Optical Society of America A*, 7(7), 1209–1217. <https://doi.org/10.1364/JOSAA.7.001209>
- National Centers for Environmental Information. (2023). *National digital forecast database*. <https://www.ncei.noaa.gov/products/weather-climate-models/national-digital-forecast-database>

- Neider, M. B., & Zelinsky, G. J. (2011). Cutting through the clutter: Searching for targets in evolving complex scenes. *Journal of Vision*, 11(14), Article 7. <https://doi.org/10.1167/11.14.7>
- Ognjanovic, S., Thüring, M., Murphy, R. O., & Hölscher, C. (2019). Display clutter and its effects on visual attention distribution and financial risk judgment. *Applied Ergonomics*, 80, 168–174. <https://doi.org/10.1016/j.apergo.2019.05.008>
- onX. (2023). *Apps for outdoor adventures*. <https://www.onxmaps.com/>
- Pankok, C., Jr., & Kaber, D. B. (2018). An integrated measure of display clutter based on feature content, user knowledge and attention allocation factors. *Ergonomics*, 61(5), 682–696. <https://doi.org/10.1080/00140139.2017.1401669>
- Poole, C. A., Warden, A. C., Wickens, C. D., Raikwar, A., Clegg, B. A., & Ortega, F. R. (2023). Information access costs with a wide-angle desktop display. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 67(1), 947–952. https://scholar.google.com/citations?view_op=view_citation&hl=en&user=cj1HVAAoAAAAJ&citation_for_view=cj1HVAAoAAAAJ:u-x6o8ySG0sC
- Psychology Software Tools. (2020). *E-prime go*. <https://support.pstnet.com/>
- Rashid, B., Zahid, F., Sohail, A., Nadeem, H. A., Faridi, T. A., Jan, U., & Hameed, S. (2021). Comparing the validity of electronic color vision test with conventional ishihara pseudoisochromatic plate test: Validity of electronic color vision test. *Pakistan BioMedical Journal*, 4(2), 156–159. <https://doi.org/10.54393/pbmj.v4i2.96>
- Rosenholtz, R., Li, Y., Mansfield, J., & Jin, Z. (2005). *Feature congestion: A measure of display clutter* [Conference session]. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. <https://doi.org/10.1145/1054972.1055078>
- Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, 7(2), Article 17. <https://doi.org/10.1167/7.2.17>
- Schons, V., & Wickens, C. D. (1993). *Visual separation and information access in aircraft display layout* (Technical Report ARL-93-7/NASA-A3I-93-1). University of Illinois, Aviation Research Laboratory.
- Pachella, R. G. (1974). The interpretation of reaction time in information-processing research 1. In B. Kantowitz (Ed.), *Human information processing* (pp. 41–82). Routledge. <https://doi.org/10.4324/9781003176688-2>
- Treisman, A., & Paterson, R. (1984). Emergent features, attention, and object perception. *Journal of Experimental Psychology: Human Perception and Performance*, 10(1), 12–31. <https://doi.org/10.1037/0096-1523.10.1.12>
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136. [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)
- Tullis, T. S. (1983). The formatting of alphanumeric displays: A review and analysis. *Human Factors*, 25(6), 657–682. <https://doi.org/10.1177/001872088302500604>
- Tullis, T. S. (1986). A system for evaluating screen formats. *Proceedings of the Human Factors Society Annual Meeting*, 30(12), 1216–1220. <https://doi.org/10.1177/1541931286030012>
- Van Staden, D., Mahomed, F. N., Govender, S., Lengisi, L., Singh, B., & Aboobaker, O. (2018). Comparing the validity of an online Ishihara colour vision test to the traditional Ishihara handbook in a South African university population. *African Vision and Eye Health*, 77(1), Article a370. <https://doi.org/10.4102/aveh.v77i1.370>
- Ververs, P. M., & Wickens, C. D. (1998). Head-up displays: Effects of clutter, display intensity, and display location on pilot performance. *The International Journal of Aviation Psychology*, 8(4), 377–403. https://doi.org/10.1207/s15327108ijap0804_4
- Vincow, M. A., & Wickens, C. D. (1993). Spatial layout of displayed information: Three steps toward developing quantitative models. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 37(4), 348–352. <https://doi.org/10.1177/154193129303700420>
- Warden, A. C., Wickens, C. D., Clegg, B. A., & Ortega, F. R. (2023). A meta-analytic examination of the cost and benefits of overlay versus separate displays. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 67(1), 1501–1502. <https://doi.org/10.1177/21695067231192430>
- Warden, A. C., Wickens, C. D., Clegg, B. A., Rehberg, D., & Ortega, F. R. (2023). Information access effort: The role of head movements for information presented at increasing eccentricity on flat panel and head-mounted displays. *Human Factors*. Advance online publication. <https://doi.org/10.1177/00187208231204567>
- Warden, A. C., Wickens, C. D., Rehberg, D., Clegg, B. A., & Ortega, F. R. (2022). Information access effort: Are head movements “cheap” or even “free”? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 66(1), 2203–2207. <https://doi.org/10.1177/1071181322661127>
- Weintrub, D., & Ensing, M. (1992). *Human factors issues in head up display design* (State of the Art Report 92-2). Wright Patterson Air Force Base Crew System Ergonomics Information Analysis Center.
- Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, 143(5), 2020–2045. <https://doi.org/10.1037/xge0000014>
- Wickens, C. D. (1993). Cognitive factors in display design. *Journal of the Washington Academy of Sciences*, 83(4), 179–201. <http://www.jstor.org/stable/24531392>
- Wickens, C. D. (2014). Effort in human factors performance and decision making. *Human Factors*, 56(8), 1329–1336. <https://doi.org/10.1177/0018720814558419>
- Wickens, C. D., Alexander, A. L., Ambinder, M. S., & Martens, M. (2004). The role of highlighting in visual search through maps. *Spatial Vision*, 17(4–5), 373–388. <https://doi.org/10.1177/154193120404801622>
- Wickens, C. D., & Andre, A. D. (1990). Proximity compatibility and information display: Effects of color, space, and object display on information integration. *Human Factors*, 32(1), 61–77. <https://doi.org/10.1177/001872089003200105>
- 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., McCarley, J., & Gutzwiller, R. (2022). *Applied attention theory* (2nd ed.). Taylor & Francis. <https://doi.org/10.1201/9781003081579>
- Wickens, C. D., Vincow, M. A., Schopper, A. W., & Lincoln, J. E. (1997). *Computational models of human performance in the design and layout of controls and displays* (CSERIAC SOAR Report 97-22). Wright Patterson Air Force Base Crew System Ergonomics Information Analysis Center.
- Wickens, C. D., & Ward, J. (2017). Cockpit displays of traffic and weather information: Effects of 3D perspective versus 2D coplanar rendering and database integration. *The International Journal of Aerospace Psychology*, 27(1–2), 44–56. <https://doi.org/10.1080/10508414.2017.1366270>
- Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review*, 28(4), 1060–1092. <https://doi.org/10.3758/s13423-020-01859-9>
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5(6), 495–501. <https://doi.org/10.1038/nrn1411>
- Yang, J. X., Park, T., Wickens, C. D., & Helander, M. G. (2014). Effects of information access cost, confidence judgment and overconfidence bias on information retrieval strategy and task performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1), 864–868. <https://doi.org/10.1177/1541931214581182>

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. (2001). Attentional filtering in the design of electronic map displays: A comparison of color coding, intensity coding,

and decluttering techniques. *Human Factors, 43*(4), 543–562. <https://doi.org/10.1518/001872001775870359>

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>

Appendix A

Computational Clutter Analyses of Experiment 1

In **Table A1**, we summarize the response time effects for the overlay display (top) and the adjacent display (bottom) for each component of the analyses: clutter domain, question type, and their interaction. The primary component of Hypothesis 3 is addressed by the range of effect sizes (see Cohen’s *d* and *t* values) for both near- and far-domain clutter for the focused attention tasks, as well as for the integration

tasks when the displays are overlaid and adjacent. The average Cohen’s *d* effect size for the effect of clutter in the overlay display condition, regardless of question type, is 0.24. In contrast, the average Cohen’s *d* effect size is 0.06 when the displays are adjacent. Taken together, the larger effect of clutter in the overlay display condition suggests that overlay clutter amplifies the effects of numerosity clutter.

Table A1
Summary of Response Time Effect Size Results for the Overlay Display Condition

| Clutter domain | Clutter effect | Question domain | Interaction |
|-------------------|------------------|-----------------|-----------------|
| Overlay display | | | |
| Focused attention | | | |
| Near | <i>d</i> = 0.22 | <i>d</i> = 0.11 | <i>d</i> = 0.12 |
| Far | <i>d</i> = 0.43 | <i>d</i> = 0.13 | <i>d</i> = 0.13 |
| Integration | | | |
| Near | <i>d</i> = 0.25 | | |
| Far | <i>d</i> = 0.07 | | |
| Adjacent display | | | |
| Focused attention | | | |
| Near | <i>d</i> = 0.04 | <i>d</i> = 1.01 | <i>d</i> = 0.39 |
| Far | <i>d</i> = 0.004 | <i>d</i> = 0.31 | <i>d</i> = 0.04 |
| Integration | | | |
| Near | <i>d</i> = 0.18 | | |
| Far | <i>d</i> = 0.02 | | |

Appendix B

Computational Clutter Analyses of Experiment 2

Table B1 presents the additional response time results from the overlay, adjacent, and separate display conditions. **Table B2** presents the additional error results in the same format. In each case, we present the effects of clutter, question domain, and the interaction between the two independent variables. The most important takeaway message from **Table B1** is the general replication

of the pattern of response time effects from Experiment 1 (compared to **Table A1**). The most important message from the error effects in **Table B2** is, again, the general replication, but the negative slope of the clutter effect on error—an improved accuracy with more clutter, again replicating this somewhat unexpected finding from Experiment 1.

(Appendices continue)

Table B1

Summary of Response Time Effect Size Results for the Overlay, Adjacent, and Separate Display Conditions

| Clutter domain | Clutter effect | Question domain | Interaction |
|-------------------|----------------|-----------------|-------------|
| Overlay display | | | |
| Focused attention | | | |
| Near | $d = 0.22$ | $d = 0.59$ | $d = 0.24$ |
| Far | $d = 0.03$ | $d = 0.21$ | $d = 0.04$ |
| Integration | | | |
| Near | $d = 0.17$ | | |
| Far | $d = 0.02$ | | |
| Adjacent display | | | |
| Focused attention | | | |
| Near | $d = 0.07$ | $d = 0.77$ | $d = 0.27$ |
| Far | $d = 0.001$ | $d = 0.61$ | $d = 0.10$ |
| Integration | | | |
| Near | $d = 0.28$ | | |
| Far | $d = 0.03$ | | |
| Separate display | | | |
| Focused attention | | | |
| Near | $d = 0.04$ | $d = 0.35$ | $d = 0.01$ |
| Far | $d = 0.15$ | $d = 0.16$ | $d = 0.11$ |
| Integration | | | |
| Near | $d = 0.32$ | | |
| Far | $d = 0.03$ | | |

Table B2

Summary of Error Rate Effect Size Results for the Overlay, Adjacent, and Separate Display Conditions

| Clutter domain | Clutter effect | Question domain | Interaction |
|-------------------|----------------|-----------------|--------------|
| Overlay display | | | |
| Focused attention | | | |
| Near | $z = -3.10$ | $z = -0.111$ | $z = -0.314$ |
| Far | $z = -5.27$ | $z = 1.76$ | $z = -1.97$ |
| Integration | | | |
| Near | $z = 4.26$ | | |
| Far | $z = 2.76$ | | |
| Adjacent display | | | |
| Focused attention | | | |
| Near | $z = -1.44$ | $z = -2.83$ | $z = 3.88$ |
| Far | $z = -1.68$ | $z = -1.68$ | $z = 2.77$ |
| Integration | | | |
| Near | $z = -8.38$ | | |
| Far | $z = -6.66$ | | |
| Separate display | | | |
| Focused attention | | | |
| Near | $z = -1.47$ | $z = -3.16$ | $z = 2.95$ |
| Far | $z = 1.88$ | $z = 1.13$ | $z = -1.48$ |
| Integration | | | |
| Near | $z = 3.47$ | | |
| Far | $z = 1.31$ | | |

Received January 23, 2023
 Revision received January 1, 2024
 Accepted January 8, 2024 ■