

## Central Lancashire Online Knowledge (CLoK)

Title	Adding depth to overlapping displays can improve visual search performance
Type	Article
URL	<a href="https://clock.uclan.ac.uk/id/eprint/22327/">https://clock.uclan.ac.uk/id/eprint/22327/</a>
DOI	<a href="https://doi.org/10.1037/xhp0000353">https://doi.org/10.1037/xhp0000353</a>
Date	2017
Citation	Godwin, Hayward J., Menneer, Tamaryn, Liversedge, Simon Paul, Cave, Kyle R., Holliman, Nick S. and Donnelly, Nick (2017) Adding depth to overlapping displays can improve visual search performance. <i>Journal of Experimental Psychology: Human Perception and Performance</i> , 43 (8). pp. 1532-1549. ISSN 0096-1523
Creators	Godwin, Hayward J., Menneer, Tamaryn, Liversedge, Simon Paul, Cave, Kyle R., Holliman, Nick S. and Donnelly, Nick

It is advisable to refer to the publisher's version if you intend to cite from the work.  
<https://doi.org/10.1037/xhp0000353>

For information about Research at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the <http://clock.uclan.ac.uk/policies/>

Running head: Depth in Visual Search

Adding Depth to Overlapping Displays can Improve Visual Search Performance

Hayward J. Godwin<sup>1</sup>, Tamaryn Menneer<sup>1</sup>, Simon P. Liversedge<sup>1</sup>, Kyle R. Cave<sup>2</sup>, Nick S.

Holliman<sup>3</sup> & Nick Donnelly<sup>1</sup>

<sup>1</sup>University of Southampton

<sup>2</sup>University of Massachusetts Amherst

<sup>3</sup>University of Newcastle

Author Note

Correspondence regarding this article should be addressed to Hayward J. Godwin, University of Southampton, School of Psychology, Highfield, Southampton, Hampshire, SO17 1BJ. Tel: +44(0)2380 595078; Email: [hayward.godwin@soton.ac.uk](mailto:hayward.godwin@soton.ac.uk).

### Abstract

Standard models of visual search have focused upon asking participants to search for a single target in displays where the objects do not overlap one another, and where the objects are presented on a single depth plane. This stands in contrast to many everyday visual searches wherein variations in overlap and depth are the norm, rather than the exception. Here, we addressed whether presenting overlapping objects on different depths planes to one another can improve search performance. Across four different experiments using different stimulus types (opaque polygons, transparent polygons, opaque real-world objects, and transparent X-ray images), we found that depth was primarily beneficial when the displays were transparent, and this benefit arose in terms of an increase in response accuracy. Although the benefit to search performance only appeared in some cases, across all stimulus types, we found evidence of marked shifts in eye-movement behavior. Our results have important implications for current models and theories of visual search, which have not yet provided detailed accounts of the effects that overlap and depth have on guidance and object identification processes. Moreover, our results show that the presence of depth information could aid real-world searches of complex, overlapping displays.

### Significance

Given the widespread availability of three-dimensional displays, we asked whether presenting displays in 3D enables searchers to find targets in overlapping displays more easily. We found that, in some cases, performance was improved when people searched through overlapping displays. The primary benefit was to displays containing overlapping transparent objects. Our results have implications for real-world tasks wherein objects need to be examined rapidly and accurately (such as airport baggage screening), but can be obscured by being overlapped by other objects.

*Acknowledgements:* All authors supported by funding from the Economic and Social Sciences Research Council (grant ref. ES/I032398/1). The authors wish to thank Kristiana Newton, Florence Greber, Stuart Pugh, and Patrick Clark for their assistance with data collection. Some of the images were provided by the Human Factors Program of the U.S. Department of Homeland Security's Transportation Security Laboratory. Note that the research reported in this report did not involve trained security screener personnel.

*Keywords:* visual search, eye movements, dual-target search, depth

Standard visual search experiments ask participants to seek out a single target within displays containing objects that are spatially separated in the plane of the monitor, such that they do not overlap with one another. In addition, the objects are typically presented on a single depth plane. Although controlled experiments of this nature have been highly valuable in studying a wide range of aspects of visual search and visual cognition (for a review, see Eckstein, 2011), they have been constrained in the sense that they have failed to tap into a number of aspects of ‘everyday’ searches. Here, we focus on two of these aspects: namely, *depth* and *overlap*. Differences in depth and overlap are the norm, rather than the exception, when we search our environment (e.g., searching for the face of a friend in a crowd, or for a set of keys in a messy office). However, surprisingly little is known with regards to how search performance is influenced by the presence of depth and overlap, and there is a significant gap in the current literature in terms of whether the presence of depth can serve to ameliorate the difficulties associated with searching through overlapping displays. At a practical level, understanding whether the presence of depth information can improve search performance is particularly important since real-world tasks, such as airport baggage screening and radiology, typically involve multiple overlapping objects. This is especially the case now given the widespread availability of displays and presentation methods involving depth.

The goal of the current set of experiments was therefore to address how the presence of depth influences visual search behavior in overlapping displays. In order to allow any benefits of adding depth to emerge, we manipulated the amount of overlap between objects in the display. At a basic level, overlap obscures diagnostic features (e.g., see Gosselin & Schyns, 2001; Schyns, 1998) that need to be used to identify objects. However, the visual information that is available at points of overlap depends on whether objects are opaque or transparent. When the objects are opaque, diagnostic features can be completely hidden, and identification must be made upon the basis of only partial information. When the objects are transparent, although the diagnostic features are still visible, they are not necessarily easy to detect and interpret due to interference from overlapping objects. Transparency results in color combinations at regions of overlap that do not necessarily match the expectations of what the target should look like (Hillstrom, Wakefield, & Scholey, 2013). Previous studies of visual search and overlap have found that increasing overlap between objects in visual search tasks increases RTs and decreases response accuracy (Bravo & Farid, 2004a, 2004b, 2006; Hillstrom et al., 2013). Here, we manipulated overlap at the level of the item and include high levels of overlap (up to 90% of an object could overlap with another). Moreover, our

approach used eye movements to determine the source of errors in responses, as explained in more detail below, which has not previously been examined in visual search of overlapping objects.

As noted above, our goal here was to determine whether the negative effects of overlap can be mitigated by separating objects in depth. With that in mind, the stimulus images in our studies were presented with different levels of three-dimensional depth. Depth has long been known to serve as a guiding feature in visual search (Finlayson, Remington, Retell, & Grove, 2013; He & Nakayama, 1992; McSorley & Findlay, 2001; Nakayama, Shimojo, & Silverman, 1989; Nakayama & Silverman, 1986; O'Toole & Walker, 1997), but it still remains relatively uncertain whether the presence of depth information can aid search performance. There have been a number of studies conducted using the presence of depth in real-world search tasks such as radiographic image screening (van Beurden, van Hoey, Hatzakis, & Ijsselstein, 2009), and these have revealed mixed results in terms of benefits to performance from adding depth (McIntire, Havig, & Geiselman, 2012, 2014). Here, we expected that the presence of depth should serve to aid in overcoming the problems associated with examining overlapping objects by acting as a cue to object segmentation processes. Half of our participants searched overlapping displays in which each object was presented on a different depth plane (the *multi-plane* condition) while the remaining participants searched displays in which all objects were presented on the same single depth plane (the *single-plane* condition).

To ensure that our results were not restricted to a particular level of search difficulty or stimulus type we used abstract objects (opaque and transparent polygons), and complex objects (opaque images of real-world household objects and transparent images from X-ray baggage screening). In addition, to assess whether our findings generalize to more difficult search tasks that are also more ecologically valid, we employed search in which either of two targets can appear (dual-target search) as well as standard single-target search. While there is evidence that searchers can successfully maintain separate search templates for both targets (Barrett & Zobay, 2014; Beck, Hollingworth, & Luck, 2012; Grubert & Eimer, 2016; Irons, Folk, & Remington, 2012; Stroud, Menneer, Cave, & Donnelly, 2012), dual-target search for two targets typically results in a *dual-target cost* (Menneer, Barrett, Phillips, Donnelly, & Cave, 2004, 2007). The dual-target cost emerges as a reduction in response accuracy, coupled with an increase in reaction times (RTs), when comparing dual-target versus single-target searches. The cost emerges both for simple, abstract colored shapes, as well as more complex objects such as those derived from airport X-ray baggage screening (Godwin, Menneer,

Cave, & Donnelly, 2010). Eye-movement experiments have revealed that the dual-target cost arises, at least in part, as a consequence of a reduction in guidance in dual-target searches (e.g., Grubert & Eimer, 2016). When searching for a single target, we fixate objects that share both visual (Becker, 2011; Luria & Strauss, 1975; Stroud, Menneer, Cave, Donnelly, & Rayner, 2011) and semantic (Godwin, Hout, & Menneer, 2014; Sobel, Puri, & Hogan, 2014) similarity with the target. In dual-target search, guidance towards targets is reduced, and searchers fixate objects that they would not have fixated in single-target searches (Menneer et al., 2012; Stroud et al., 2012). There is also evidence that the dual-target cost arises as a consequence of bottlenecks in the processing of object identities, alongside any detriments to guidance processes (Godwin, Walenchok, Hout, Hout, & Goldinger, 2015; Houtkamp & Roelfsema, 2009).

In order to better understand how and why behavioral performance varied as a function of these modulations, especially the presence of depth in displays, we recorded participants' eye movements as they searched. Eye movements are known to provide an excellent index of online cognitive processing (Liversedge & Findlay, 2000; Rayner, 2009), and here we focused on the *errors* made when searching to determine the sources of shifts in behavioral performance (Cain, Adamo, & Mitroff, 2013; Nodine & Kundel, 1987; Schwark, MacDonald, Sandry, & Dolgov, 2013). Search is often broken down into two components: *perceptual selection*, wherein target-similar objects are selected for detailed processing (this can be compared to guidance in search, see Wolfe, Cave, & Franzel, 1989); and *perceptual identification*, wherein objects are fixated and responded to. When either selection or identification processes fail, targets are missed (Cain et al., 2013; Godwin, Menneer, Riggs, Cave, & Donnelly, 2015; Godwin, Menneer, Riggs, Taunton, et al., 2015; Hout, Walenchok, Goldinger, & Wolfe, 2015; Moore & Osman, 1993; Nodine & Kundel, 1987).

Perceptual selection is measured in terms of the probability of fixating target objects, and the time taken to fixate target objects. When participants are less likely to fixate target objects (assuming that those targets need to be foveated for veridical identification), it is extremely unlikely that those targets will be detected, which can account for a reduction in response accuracy. When participants are slower to fixate targets, it can increase response time on target-present trials. On the other hand, perceptual identification is measured in terms of the probability identifying targets after fixating them, and the time taken to respond to a target once fixated (i.e., verification time: see Castelano, Pollatsek, & Cave, 2008). The probability of fixating and then identifying targets serves as a measure of the difficulty of object identification, and directly relates to response accuracy rates. Verification time

(Rayner, Smith, Malcolm, & Henderson, 2009) provides a measure of the difficulty of object identification: the longer this process, the more difficult object identification is, and the longer the overall response time.

Previous studies have used eye movements to show impairment of both perceptual selection and perceptual identification processes in dual-target search (Godwin, Walenchok, et al., 2015; Menneer et al., 2012; Stroud et al., 2012), but it remains an open question as to which of these processes underlies effects of overlap, and whether the presence of depth facilitates either of these processes. Since increasing levels of overlap obscures diagnostic features, we predicted that increasing overlap would increase both perceptual selection and perceptual identification errors. However, we anticipated these effects may be reduced for the transparent stimulus types because diagnostic features for these stimulus types are still available despite being obscured. We predicted that the presence of depth would improve behavioral performance and reduce both perceptual selection and identification errors. We expected these improvements to occur because the presence of depth should enable participants to more readily segregate overlapping objects into their constituent parts and therefore select and identify them more effectively. Generally speaking, we anticipated that the benefits from adding depth to displays would be greater for the transparent stimulus types than the opaque stimulus types. This is because the transparent stimuli are fully visible, and interpretation of ambiguity at overlap could be facilitated by separation in depth, while overlapped parts of the opaque stimuli are not visible, and thus there is no information to be enhanced by the presence of depth.

From a theoretical standpoint, the present set of experiments push forward current models of visual search (Thornton & Gilden, 2007; Wolfe, 2007), including recent models of eye movements and search (Godwin, Reichle, & Menneer, in press; Zelinsky, 2008), by considering factors not currently captured within those models such as the way that overlap interferes with object identification. Many models focus, quite rightly, on guidance of visual attention towards the target in search. Looking to the future, as research of more complex search tasks increases, models need to take into account difficulties that arise from overlap and dual-target search, as well as the potentially beneficial effects of facilitating object segmentation (here, by separation in depth). From a practical standpoint, the present set of experiments will be of benefit to real-world tasks by enabling a better understanding of how and why search can fail when searching for multiple targets in complex, overlapping displays, and in terms of whether or not the presence of depth can aid search performance in displays of this type.

To prelude our findings, we did find evidence that the presence of depth can improve search performance. Adding depth to the displays improved response accuracy for transparent stimuli, as well as for the real-world stimuli in target-present trials. However, depth did not influence RTs substantially, and this was the case across the different stimulus types. Our examinations of the eye movement data revealed that the presence of depth in the displays primarily influenced the probability that participants would fixate targets after identifying them, and also attenuated the effects of overlap for the probability of fixating targets. Unlike response accuracy, depth had these influences across all stimulus types, and was not restricted to certain stimulus types only.

## Method

We conducted four Experiments with the same basic design, with the only difference between them being the different stimulus types (Opaque Polygons, Transparent Polygons, Real-world Objects, and X-ray Objects). We present the Method for all four experiments together given their similar nature.

### Participants

Prior to participation, participants in all experiments completed a series of vision tests to ensure that they had normal color vision (Ishihara, 1964) and a sufficient degree of 3D depth perception, as assessed by the Titmus Stereo Test (i.e., a score of nine for the Wirt Circles component). All participants were postgraduates and staff from the University of Southampton, who took part either for course credits or for payment (£36). 32 participants were recruited to take part searching each stimulus type, with 128 participants in total.

### Apparatus

Eye movement behavior was recorded using an Eyelink 1000 operating at 1000Hz (i.e., 1 sample per millisecond). The experiment was implemented in SR Research Experiment Builder. Viewing of the displays was binocular, though only the right eye was recorded<sup>1</sup>. A nine-point calibration was used, and accepted only when the mean error was less than 0.5° of visual angle, with no single error exceeding 1° of visual angle. A drift

---

<sup>1</sup> We recorded monocular eye movements because we wished only to be able to identify fixation locations and their durations. We did not aim to determine the point of a given fixation within the depth plane, and consequently, binocular eye movements and vergence responses were not measured.



correct procedure was performed before each trial and calibrations were repeated when the error exceeded  $1^\circ$  of visual angle. Note that throughout the experiment, calibration points and drift corrects were presented in stereoscopic depth so as to minimize any disruption from swapping from two-dimensional displays in the plane of the monitor to stereoscopic displays. Eye movement data were parsed into fixations and saccades using the recommended default settings for the Eyelink system (i.e., saccades were detected using a velocity threshold of  $30^\circ$  per second or an acceleration that exceeded  $8000^\circ$  per second-squared).

Stimuli were presented on a Hyundai W243s monitor with a 60Hz refresh rate and a resolution of 1920 x 1200 pixels. Participants sat 86cm from the computer display in a dimly-lit room and wore a pair of polarized spectacles. Head position was stabilized using a chinrest and responses (“target-present” or “target-absent”) were made using a gamepad response box.

## Stimuli

### Display layout.

The stimuli were pre-generated prior to data collection using custom code written using C# (see Godwin, Holliman, et al., in press, for more detail). The full visual area of the computer monitor subtended  $42.8^\circ \times 26.7^\circ$  of visual angle. To avoid any depth artefacts that can occur at the edge of such large displays, we only presented search images within a central  $28.5^\circ \times 21.4^\circ$  region within the display, leaving the outer edges blank. The search arrays (see Figure 1) were set out according to a 4 by 3 grid of *master grid cells* (with each cell subtending  $7.1^\circ \times 7.1^\circ$  degrees of visual angle). Master grid cells were randomly selected to be occupied by objects. In half of the displays, 6 of the master grid cells were randomly selected to contain objects and in the other half 10 were selected. Objects within each master grid cell could not fall along a cell boundary to ensure that objects could not be immediately adjacent to, or overlap, objects in adjacent master grid cells when the displays were interlaced for 3D depth (see below ‘Implementation of Depth’ section for more detail on this process). In order to make the displays appear less systematic, occupied master grid cells were jittered by a random amount into adjacent master grid cells by a distance of up to  $3.6^\circ$  of visual angle, but only if the adjacent master grid cells were not already occupied. When it was determined that a master grid cell should be occupied, four objects were placed within that master grid cell. To begin with, the master grid cell was sub-divided into four inner grid cells

(i.e., quadrants) of equal size (i.e.,  $3.6^\circ \times 3.6^\circ$  of visual angle), each with an object contained within. When there was to be no overlap of objects, the objects remained in these positions.

### **Opaque polygons.**

The polygon distractors were four-, five- and six-sided polygons of different colors that were of a randomly generated shape. The shapes were generated as follows. The points of each polygon were assigned relative to the center point in each inner grid cell. From this center point, a range of angles from the center point were set out where each of the points could fall. These ranges were adaptive depending on the number of vertices of the polygon, and were adopted to prevent individual points overlapping with one another. After selecting an angle relative to the center point inside each inner grid cell, each vertex was then set at a randomly-determined distance from the center point. This distance was selected to range between  $1.6^\circ$  and  $3.2^\circ$  of visual angle, with the constraint that no two consecutive points could be within  $0.04^\circ$  of each other (this was in order to avoid neighboring sides of the polygons from being absorbed into a single side when their distances from the center were very similar). Once the points were selected, they were then connected, in order, to form the polygon. The same algorithm was used to generate both target and distractor objects.

Target and distractor colors were selected using the set of 16 colors used in previous studies including Menneer et al. (2007) and Stroud et al. (2012). These colors form a ring in CIExyY space that represents a wide range of different hues, and no single color's relative salience causes it to pop-out from the others. The number of distractors of each color type was controlled across the trials so that an equal number of each color was presented during the course of the experiment: however, on each trial, the specific colors of the distractors were randomly selected. Note that some distractors were of identical color to the targets.

For each participant, a pair of unique polygons was generated, using the same algorithm as the distractor polygons, and designated arbitrarily as Target A and Target B. One of the targets had four vertices, the other had six vertices, and for all participants, the two target colors were eight steps apart from each other in the color ring, making them maximally different from one another. In order to ensure that our results were applicable across different colors, participants were asked to search for different pairs of target colors selected from the color ring.

### **Transparent polygons.**

We used the same set of stimuli generated for the opaque polygons for the transparent polygon objects; the only difference was that the overlapping regions were set to appear as 'transparent'. This involved taking the combined color of overlapping regions to give the

appearance of transparency. The combined color was calculated by examining each pixel in the display and combining the red, green, and blue values for different objects when they were to overlap. This was achieved by computing the proportion of maximum in each of the three color channels (i.e., out of 255) from each object at each pixel, and then multiplying the values together. So, for example, if Object A had a red value of 100 at a given pixel, and object B had a red value of 100 at the same pixel, the resultant value would be  $((100/255) * (100/255)) * 255 = (0.39 * 0.39) * 255 =$  a combined red value of 39.

#### **Real-world objects (opaque).**

The real-world objects were taken from the Hemera photo-objects database. For targets, we selected weapons (guns/knives) and toys that were primarily blue in color. For distractors, we selected an array of household objects and items (e.g., vases, hats, bottles, binoculars, apples, etc.). In total, there were 130 objects in the weapons category and 122 objects in the toys category, with 671 objects in the distractors category. After an object was selected for presentation, it was randomly rotated by increments of 90 degrees around its original orientation.

#### **X-ray objects (transparent).**

The X-ray objects were from the same library used in previous studies (e.g., Menneer, Cave, & Donnelly, 2009). Targets were grouped into two categories: weapons (guns/knives, which both appear as blue in color) and Improvised Explosive Devices (IEDs, which appear as a mix of orange and blue in color). Distractors consisted of items normally found in luggage (e.g., coins, wallets, headphones, etc.). We used 201 weapon images, 71 IED images, and 1,303 distractor images. All X-ray images were X-ray photographs of real objects. As with the real-world objects, after an object was selected for presentation it was randomly rotated by increments of 90 degrees around its original orientation. The calculation of color at overlapping regions was determined in the same manner as for the transparent polygons.

#### **Overlapping of objects.**

Objects could only overlap with objects in the same master grid cell. When overlap was required, half of the master cells were randomly selected to contain overlapping objects, and either two or four objects within each master grid cell would overlap with one another (referred to as “maximum-layers” later, though our analyses average over this factor to focus on the core issues of interest). The *Overlap* factor was manipulated over three levels, No Overlap (0%), Medium Overlap (45%) and High Overlap (90%). The percentages refer to the percent of each object’s area that could, at maximum, overlap with any other object. Given that the precise percentage of overlap depended on the object shape and area, the exact

percentage could not always be achieved but was approximated as closely as possible without being exceeded. The target overlapped with a distractor on half of the target-present trials only, and did not overlap with distractors on the other half. This control was implemented in order to prevent participants learning that the target would fall in a cluster of overlapping objects.

### **Implementation of depth.**

Depth was implemented across four possible depth planes, with two planes appearing in front of the monitor and two planes behind the monitor. The total perceived depth range for the monitor was approximately 12.5 cm. The layers were equidistant in depth from one another. In order to implement stereoscopic depth, the image was first created for each trial, after which alternate rows of pixels within each image were transposed to the left or right. When viewed on a 3D monitor with polarized glasses, alternate pixel rows are visible to separate eyes, thereby creating the perception of depth in the images.

Two different *Depth* conditions were utilized. In the *single-plane* condition all objects were assigned to the same depth plane within a given trial. For the single-plane condition, each of the four possible depth planes was equally likely to be chosen for a trial. In the *multi-plane* condition, each of the four objects within each master grid cell was randomly assigned to a different depth plane. In the *multi-plane* condition, the target appeared at each of the four depths an equal number of times in order to prevent participants from learning that the target appeared in given layers. This manipulation meant that an overlapping target could appear at the top of the pile of overlapping objects, and therefore, be unoccluded. As will be seen in the Results, these targets were removed from the target-focused analyses<sup>2</sup>.

Finally, it is important to note that object size was not adjusted as a function of depth. However, the real-world object and X-ray images were not to scale, and the polygons were unfamiliar shapes that would not have a familiar and predictable size for participants. The displays, therefore, did not appear any more artificial in the multi-plane than the single-plane condition.

---

<sup>2</sup> Even though an overlapping target was unoccluded when it appeared in the nearest depth layer, it differed from unoccluded non-overlapping targets due to crowding from overlapping objects.

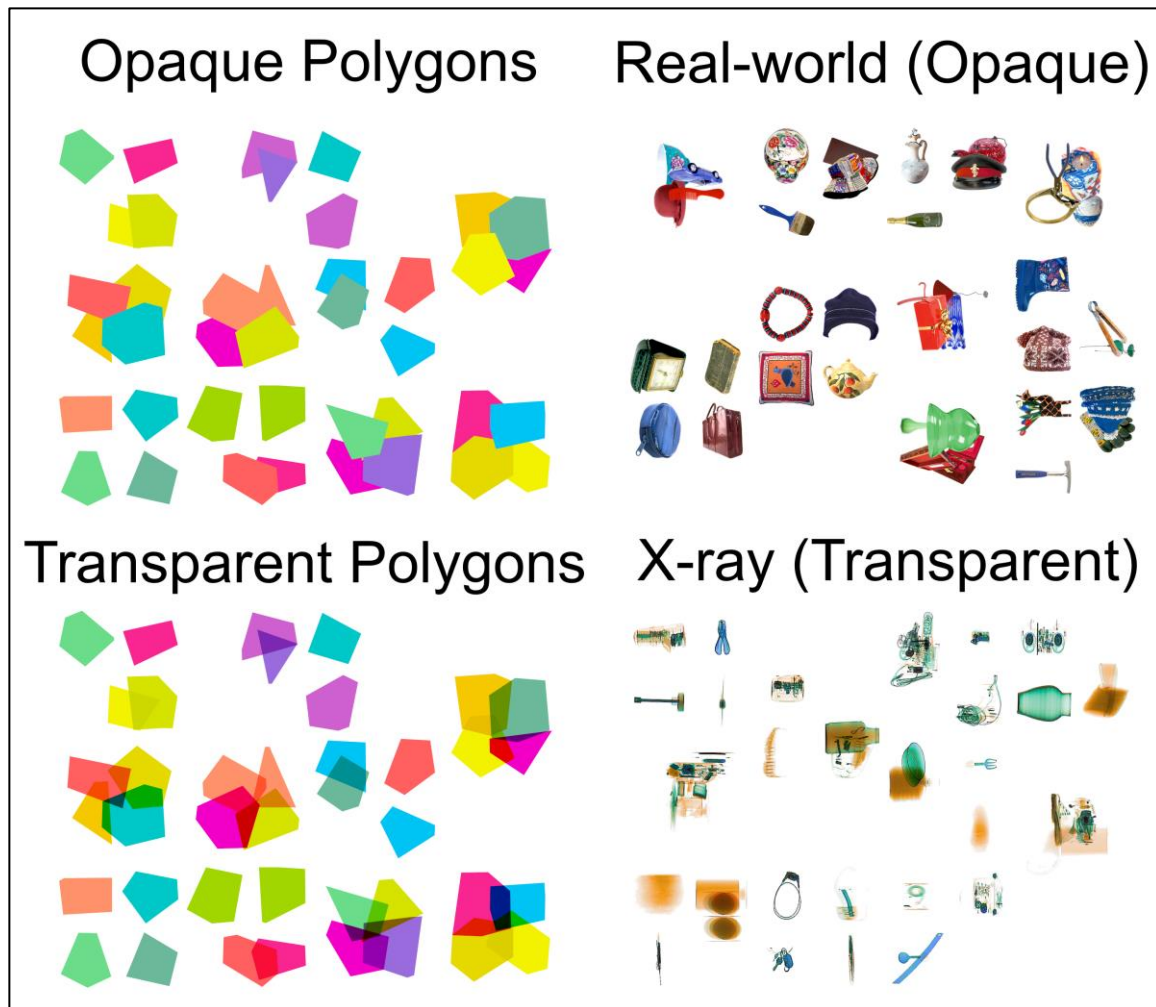


Figure 1. Examples of the different stimulus types including overlap.

### Design and Procedure

Each participant was involved in a series of six testing sessions, comprising two sessions of search at each level of the Overlap factor (0%, 45%, 90%). In half of the sessions, the overlapping of objects was controlled such that only two objects could overlap each other at any one point; in the remaining half of the sessions, up to four objects could overlap one another at any one point. This was implemented in order to vary the complexity of the displays, and to mirror real-world searches in which many objects could overlap. For the purposes of our analyses, we collapsed across the different sessions since this factor was not of primary interest. Participants could take part in no more than two testing sessions in a single day, with at least a one-hour break between sessions. The Depth factor was manipulated between participants to prevent any crossover or interference that could occur from participants searching in the multi-plane and single-plane displays.

Each testing session lasted up to 90 minutes and consisted of three blocks of 96 trials, each preceded by 16 practice trials. Participants were asked to search for two targets, which remained constant for all of their sessions. During each session, participants spent one block searching for Target A, one block searching for Target B, and one block searching for both targets A and B (dual-target search). The order of these blocks was counterbalanced across participants, though each participant experienced the same block ordered through all of their sessions. The participant's task was to respond as to whether a target was absent or present in each display. In dual-target search, only one target was present in the display at most.

A single target was presented on 48 trials in each block (i.e., 50% of trials), with both targets presented an equal number of times in the dual-target condition. Each trial began with a drift correct procedure, after which participants were presented with a reminder of the target(s) at the center of the display that they had to fixate for 500ms for the trial to begin. The reminder was presented in stereoscopic depth at the second depth layer to avoid switching costs between 2D and 3D. Following an incorrect response, a tone sounded to notify participants of their error.

## Results

### Analytic Approach

The raw data were processed using the *eyeTrackR* (Godwin, 2013) package for R (R Development Core Team, 2013). Analyses were performed using the *ez* R package (Lawrence, 2015). We examined two behavioral measures (response accuracy and response times), and four eye movement measures (time to fixate targets, probability of fixating targets, verification time, and the probability of identifying targets after fixating them). The eye movement measures were selected in order to understand whether participants exhibited failures in guidance and/or object identification under the different levels of depth and overlap. To reduce skew in the data, all time-based measures were log-transformed prior to analyses, and all proportion-based measures were arcsine-square-root transformed prior to analyses (though we report raw means in the figures).

For each of the measures, we began by conducting an initial mixed-design ANOVA. For all measures, the initial ANOVAs contained factors of Depth (single-plane, multi-plane), Search Type (single-target, dual-target), Overlap (0%, 45%, 90%) and Stimulus Type (opaque polygons, transparent polygons, real-world, X-ray). For the behavioral measures, we added the factor of Presence (present, absent) to compare target-absent and target-present

trials. The eye-movement measures require a target to be present, so were conducted on target-present trials only. Data within each session were averaged across the different stimulus set sizes and the different numbers of objects (two or four) that could overlap.

The primary goal of these analyses was to determine similarities and differences in the results across stimulus types (Opaque Polygons, Transparent Polygons, Real-world Objects, and X-ray Objects) and whether these are influenced by presenting objects on different depth planes to one another. Therefore, where appropriate, further ANOVAs were conducted focused on the different stimulus types and depths. All significant interactions were explored using Bonferroni-corrected *t*-tests. We report generalized eta-squared ( $\eta^2$ ) as a measure of effect size in the ANOVAs (Bakeman, 2005).

We removed any trials with a response time of less than 200ms as outliers (this resulted in the removal of 65 trials). Due to software and hardware errors, there was a small degree of data loss (a total of 0.08% of trials). However, no participants were left with empty cells for analysis within any of the measures. The final dataset comprised data from 220,936 search trials.

## Visualization of Results

Given the number of factors involved in analyzing the different measures below (i.e., involving the possibility of five-way interactions), for each measure, we present not only the raw descriptive statistics, but also, where relevant, the effects that depth has upon behavior and performance. Each figure begins with the ‘Overall’ results for each measure, and then additionally involves a depiction of the higher-order interactions relating to depth, aggregated across any factors that were not involved in those interactions. The goal in doing so is to more readily visualize the effects that the primary factor of interest—depth—has upon the different measures. Where depth was non-significant (or where depth was significant but post-hoc tests failed to reveal any effects), we did not include these additional figures.

## Response Accuracy

The initial ANOVA conducted on the response accuracy rates revealed main effects for all factors, plus a number of interactions (see Table 1 for ANOVA results; see Figure 2 for means). Overall, response accuracy was higher in multi-plane than single-plane displays, higher on target-absent trials than target-present trials, lower in dual-target search than single-target search, and reduced as overlap increased. The main effect of Depth was qualified by significant interactions. Despite the large number of interactions, they were subsumed within

two significant four-way interactions: namely, Depth x Stimulus Type x Overlap x Presence, and Stimulus Type x Search Type x Overlap x Presence. These two interactions were then examined in turn.

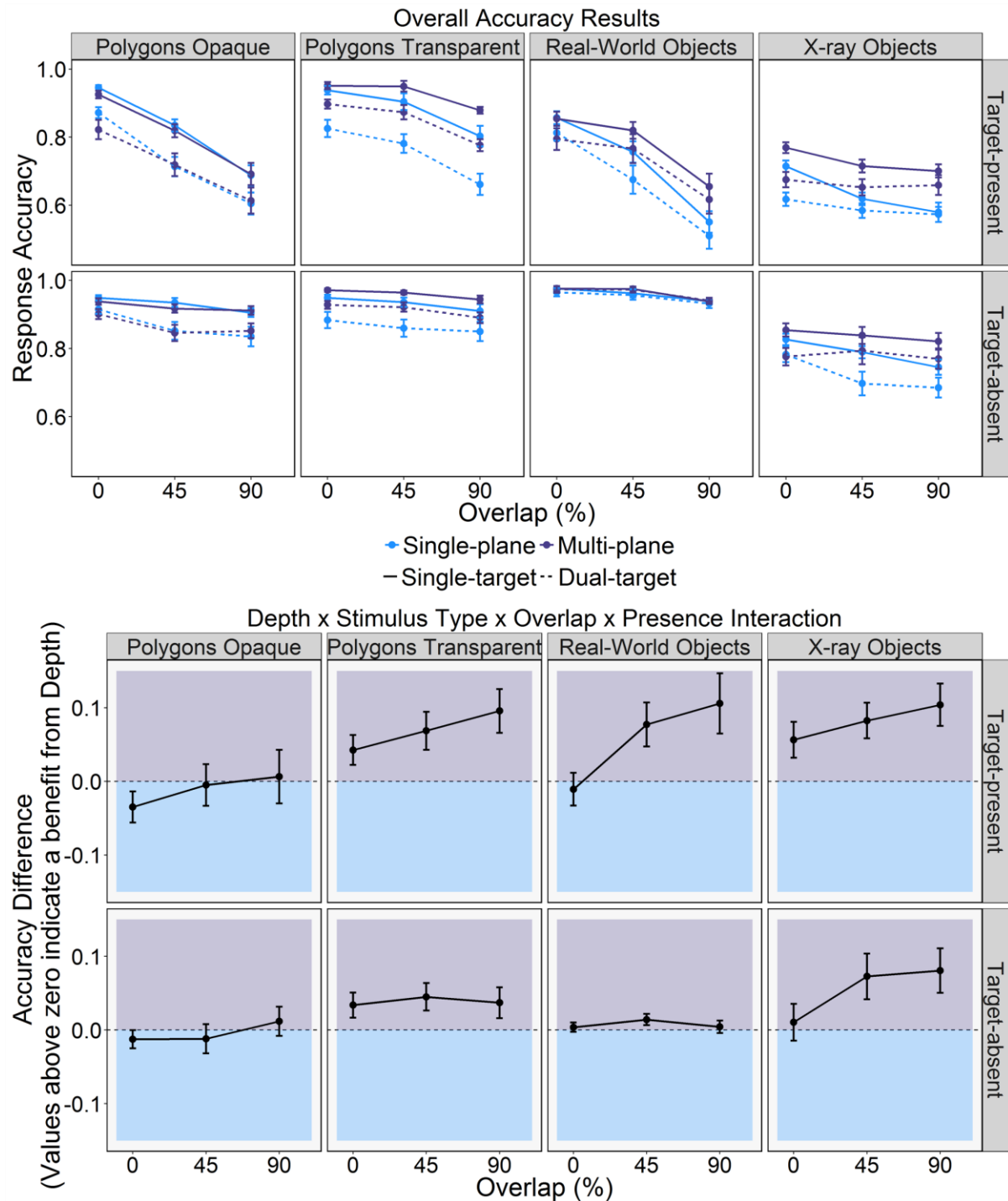


Figure 2. Overall Response Accuracy rates (upper panel) and the mean Accuracy Difference between Multi-plane and Single-plane Search (lower panel), for the different Stimulus Types,



levels of Depth, Search Types, and Target-present and Target-absent Trials. Error bars represent SE.

Table 1.

*Main Effects and Interactions for the Initial ANOVA for Response Accuracy.*

Effect/Interaction	<i>F</i>	<i>df</i>	<i>ges</i>
Depth	13.5***	(1,120)	0.04
Stimulus Type	51.33***	(3,120)	0.33
Search Type	399.86***	(1,120)	0.11
Overlap	312.55***	(2,240)	0.20
Presence	266.24***	(1,120)	0.38
Depth x Stimulus Type	2.73*	(3,120)	0.03
Depth x Search Type	2.29	(1,120)	0.001
Stimulus Type x Search Type	14.61***	(3,120)	0.01
Depth x Overlap	14.4***	(2,240)	0.01
Stimulus Type x Overlap	16.38***	(6,240)	0.04
Depth x Presence	2.15	(1,120)	0.01
Stimulus Type x Presence	17.9***	(3,120)	0.11
Search Type x Overlap	2.57	(2,240)	0.001
Search Type x Presence	18.59***	(1,120)	0.01
Overlap x Presence	134.8***	(2,240)	0.08
Depth x Stimulus Type x Search Type	2.11	(3,120)	0.002
Depth x Stimulus Type x Overlap	0.67	(6,240)	0.002
Depth x Stimulus Type x Presence	0.48	(3,120)	0.003
Depth x Search Type x Overlap	1.94	(2,240)	0.001
Stimulus Type x Search Type x Overlap	4.51***	(6,240)	0.004
Depth x Search Type x Presence	0.03	(1,120)	0.0000
Stimulus Type x Search Type x Presence	3.14*	(3,120)	0.004
Depth x Overlap x Presence	3.34*	(2,240)	0.002
Stimulus Type x Overlap x Presence	19.68***	(6,240)	0.04
Search Type x Overlap x Presence	1.33	(2,240)	0.0005
Depth x Stimulus Type x Search Type x Overlap	0.45	(6,240)	0.0004
Depth x Stimulus Type x Search Type x Presence	0.65	(3,120)	0.001
Depth x Stimulus Type x Overlap x Presence	2.14*	(6,240)	0.004
Depth x Search Type x Overlap x Presence	0.21	(2,240)	0.0001
Stimulus Type x Search Type x Overlap x Presence	2.18*	(6,240)	0.002
Depth x Stimulus Type x Search Type x Overlap x Presence	1.3	(6,240)	0.001

*Note:* \*= $p < .05$ , \*\*= $p < .01$ , \*\*\*= $p < .001$

We examined the Depth x Stimulus Type x Overlap x Presence interaction by analyzing each Stimulus Type separately, and focused on the effect of Depth. For Opaque Polygons, there were no effects or interactions relating to Depth ( $F_s < 1.2$ ). For Transparent Polygons, however, there was a main effect of Depth ( $F(1,30)=8.61, p=.006, \eta^2=.12$ ), with response accuracy being higher for multi-plane than single-plane displays. For Real-world objects the interaction between Depth x Overlap x Presence reached significance ( $F(2,60)=8.03, p<.01, \eta^2=0.02$ ). This interaction for Real-world objects arose because there was no effect of Depth for absent trials ( $F < 1$ ), while for present trials, Overlap had less of an effect in multi-plane than single-plane displays (multi-plane:  $F(2,30)=77.64, p<.001, \eta^2=0.27$ ); single-plane:  $F(2,30)=91.26, p<.001, \eta^2=0.51$ ). For X-ray stimuli, there was a main effect of Depth ( $F(1,30)=16.69, p<.001, \eta^2=0.12$ ), which interacted with Overlap ( $F(2,60)=9.65, p<.001, \eta^2=.02$ ). Response accuracy in multi-plane X-ray displays was higher than for single-plane displays in the 45% and 90% Overlap conditions only ( $t_s > 3.8, p_s < .001$ ). It therefore appears that depth primarily improves response accuracy for transparent displays, as predicted.

The Stimulus Type x Search Type x Overlap x Presence interaction was examined in the same way as the previous interaction. The Opaque Polygons, Transparent Polygons, and Real-world stimuli showed no evidence of Search Type x Overlap x Presence interactions ( $F_s < 1.9, p_s > .4$ ), yet all showed evidence of effects of Presence, Search Type and Overlap ( $F_s > 31, p_s < .001$ ). For the X-ray stimuli, there was a significant Search Type x Overlap x Presence interaction ( $F(2,60)=3.38, p=.041, \eta^2=.006$ ). This interaction arose because there was a dual-target cost for absent trials in the X-ray stimuli ( $F(1,30)=34.28, p<.001, \eta^2=.08$ ), though on present trials, the magnitude of the cost varied across levels of Overlap ( $t_s > 6, p_s < .001$ ). Generally speaking, we therefore found evidence of a standard dual-target cost for the response accuracy data.

In summary, accuracy was higher on absent than present trials, and was higher in single- than dual-target search. Accuracy also reduced with increasing overlap. Importantly, adding depth to displays was beneficial when objects in the displays were transparent (transparent polygons and X-ray stimuli). These results are in line with our predictions, and in terms of the dual-target cost, are in line with our previous experiments.

## Response Times

As with the response accuracy rates, the initial ANOVA conducted upon the (correct-response) RTs revealed a number of main effects, as well as a number of interactions (see

Table 2 for ANOVA results; see Figure 3 for means). Overall, RTs were longer in target-absent than target-present trials, were longer in dual-target than single-target search, and increased as overlap increased. Unlike with the response accuracy analyses, there was no main effect of Depth ( $F < 1$ ). The interactions that reached significance for the RTs were subsumed as part of two core four-way interactions: Depth x Search Type x Overlap x Presence, and Stimulus Type x Search Type x Overlap x Presence. We examined each of these interactions in turn.

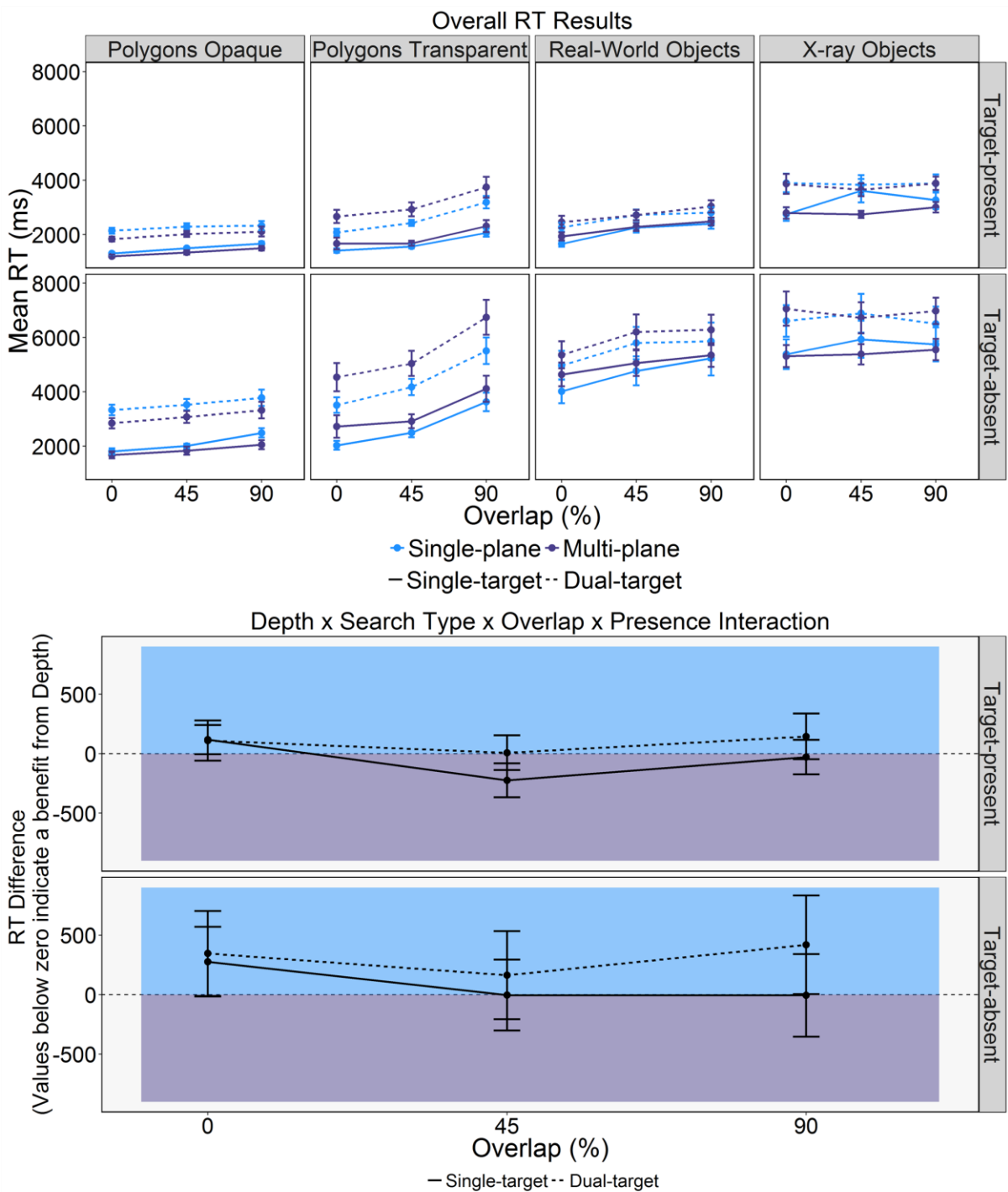


Figure 3. Mean Reaction Times (upper panel) and the Mean RT Difference between Multi-plane and Single-plane Search (lower panel) for the different Stimulus Types, levels of Depth, Search Types, and Target-present and Target-absent Trials. Error bars represent SE.

Table 2.

Main Effects and Interactions for the Initial ANOVA for Response Times.

Effect/Interaction	<i>F</i>	<i>df</i>	<i>ges</i>
--------------------	----------	-----------	------------

Depth	0.11	(1,120)	0.001
Stimulus Type	35.14***	(3,120)	0.39
Search Type	1064.26***	(1,120)	0.22
Overlap	121.26***	(2,240)	0.10
Presence	1417.48***	(1,120)	0.49
Depth x Stimulus Type	1.35	(3,120)	0.02
Depth x Search Type	1.49	(1,120)	0.0004
Stimulus Type x Search Type	50.52***	(3,120)	0.04
Depth x Overlap	2.23	(2,240)	0.002
Stimulus Type x Overlap	12.1***	(6,240)	0.03
Depth x Presence	1.41	(1,120)	0.001
Stimulus Type x Presence	35.73***	(3,120)	0.07
Search Type x Overlap	17.07***	(2,240)	0.003
Search Type x Presence	69.93***	(1,120)	0.004
Overlap x Presence	37.32***	(2,240)	0.003
Depth x Stimulus Type x Search Type	1.88	(3,120)	0.002
Depth x Stimulus Type x Overlap	0.28	(6,240)	0.001
Depth x Stimulus Type x Presence	0.6	(3,120)	0.001
Depth x Search Type x Overlap	2.86	(2,240)	0.0004
Stimulus Type x Search Type x Overlap	2.15*	(6,240)	0.001
Depth x Search Type x Presence	1.41	(1,120)	0.0001
Stimulus Type x Search Type x Presence	36.33***	(3,120)	0.01
Depth x Overlap x Presence	3.06*	(2,240)	0.0003
Stimulus Type x Overlap x Presence	15.38***	(6,240)	0.004
Search Type x Overlap x Presence	1.57	(2,240)	0.0001
Depth x Stimulus Type x Search Type x Overlap	0.34	(6,240)	0.0001
Depth x Stimulus Type x Search Type x Presence	0.67	(3,120)	0.0001
Depth x Stimulus Type x Overlap x Presence	0.27	(6,240)	0.0001
Depth x Search Type x Overlap x Presence	4.76**	(2,240)	0.0003
Stimulus Type x Search Type x Overlap x Presence	4.27***	(6,240)	0.001
Depth x Stimulus Type x Search Type x Overlap x Presence	1.4	(6,240)	0.0002

Note: \*= $p < .05$ , \*\*= $p < .01$ , \*\*\*= $p < .001$

We began by examining the Depth x Search Type x Overlap x Presence interaction, by examining present and absent trials separately (see Figure 3, lower panel). The Depth x Search type x Overlap interaction only reached significance in absent trials ( $F(2,240)=5.52$ ,  $p=.005$ ,  $ges=0.001$ ; present trials:  $F(2,240)=2.7$ ,  $p=.07$ ,  $ges=0.001$ ). Surprisingly, depth did not significantly change RTs for single-target or dual-target searches at any level of overlap when the target was absent as a main effect ( $F_s < 1$ ). The interaction was instead caused by a

larger dual-target cost in multi-plane than single-plane displays when there was 90% overlap and the target was absent (multi-plane:  $F(1,60)=331.8$ ,  $p<.0001$ ,  $ges=0.22$  ; single-plane:  $F(1,60)=238.2$ ,  $p<.0001$ ,  $ges=0.13$ ).

Next, we examined the Stimulus Type x Search Type x Overlap x Presence interaction. As with the accuracy analyses, we examined each stimulus type separately with further ANOVAs. For the Opaque Polygons, there was an interaction between Search Type and Overlap ( $F(2,60)=12.12$ ,  $p<.0001$ ,  $ges=0.02$ ), with a dual-target cost present at all levels of overlap which reduced marginally as overlap increased ( $ts>14$ ,  $ps<.0001$ ). For the other stimulus types, there were interactions between Search Type, Overlap, and Presence ( $F_s>4$ ,  $ps<.05$ ). A dual-target cost was found in all conditions, but the magnitude of this cost varied across presence and overlap ( $ts>4$ ,  $ps<.0001$ ).

In summary, the evidence shows RTs to be longer on absent than present trials, in dual- than single-target search, and to increase with overlap. Depth had little effect on RTs, other than to result in a larger dual-target cost in 90% overlap target-absent trials.

### **Eye Movement Measures: Examining Failures of Perceptual Selection and Perceptual Identification**

Taken together, the analyses of the behavioral measures indicated that depth did aid search performance by increasing response accuracy, though response times were also increased. Our analyses of the eye movement measures focused upon the types of errors made during search, in terms of failures of perceptual selection (time to fixate targets, probability of fixating targets) and perceptual identification (verification time, probability of identifying targets after fixating them).

#### **Time to fixate targets.**

The time taken to fixate targets serves as a measure of failure of perceptual selection in search. For this measure, the initial ANOVA (see Table 3 for ANOVA results and Figure 4 for means) revealed that participants required more time to fixate targets in dual-target search than single-target search, and required more time to fixate targets as overlap increased. There was no main effect of Depth ( $F<1$ ). There were a number of interactions, which were subsumed as part of a three-way interaction between Stimulus Type x Search Type x Overlap.

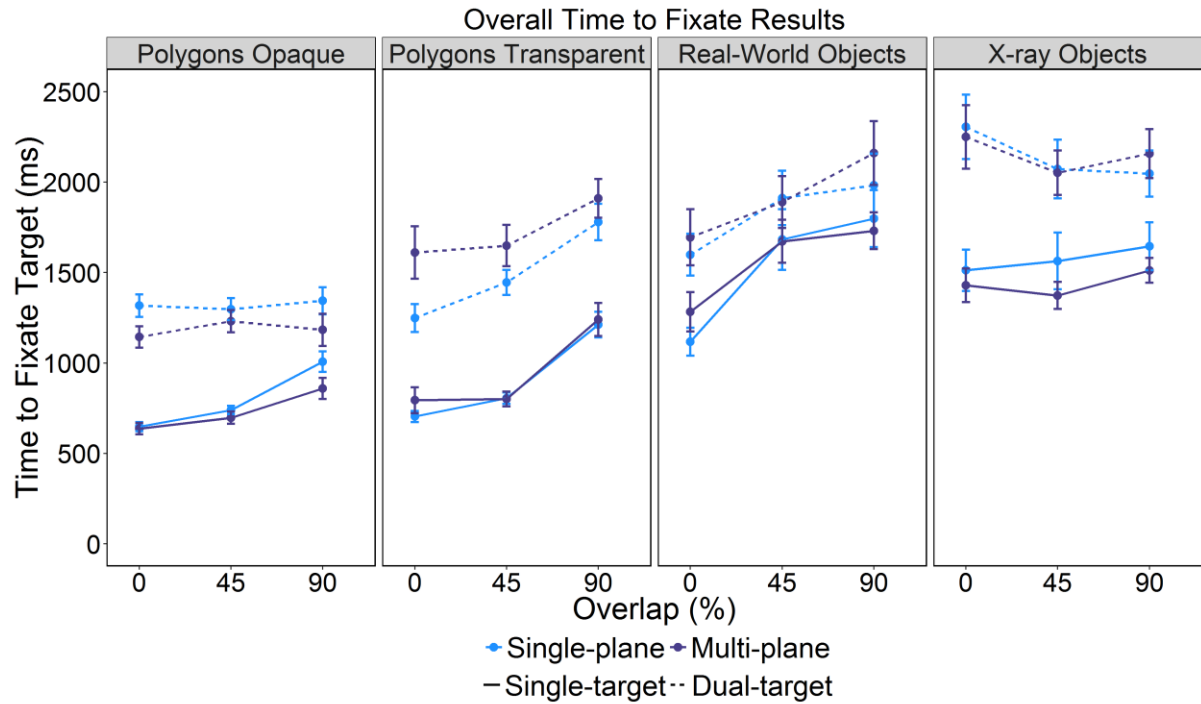


Figure 4. Time to Fixate Targets for the different Stimulus Types, levels of Depth and Search Types. Error bars represent SE.

Table 3.

*Main Effects and Interactions for the Initial ANOVA for Time to Fixate Target.*

Effect/Interaction	<i>F</i>	<i>df</i>	<i>ges</i>
Depth	0	(1,120)	0.0000
Stimulus Type	39.2***	(3,120)	0.41
Search Type	1551.36***	(1,120)	0.42
Overlap	74.9***	(2,240)	0.09
Depth x Stimulus Type	0.86	(3,120)	0.01
Depth x Search Type	3.64	(1,120)	0.002
Stimulus Type x Search Type	49.02***	(3,120)	0.07
Depth x Overlap	1.02	(2,240)	0.001
Stimulus Type x Overlap	17.61***	(6,240)	0.07
Search Type x Overlap	51.64***	(2,240)	0.03
Depth x Stimulus Type x Search Type	1.65	(3,120)	0.002
Depth x Stimulus Type x Overlap	1.22	(6,240)	0.005
Depth x Search Type x Overlap	1.61	(2,240)	0.001
Stimulus Type x Search Type x Overlap	7.2***	(6,240)	0.01
Depth x Stimulus Type x Search Type x Overlap	1.59	(6,240)	0.003

Note: \*= $p < .05$ , \*\*= $p < .01$ , \*\*\*= $p < .001$

As with the other measures, to break down the Stimulus x Search Type x Overlap interaction, we analyzed each Stimulus Type separately. The interaction between Search Type and Overlap was significant for all stimulus types ( $F_s > 3$ ,  $p_s < .04$ ). For Transparent Polygons and Real-world objects, overlap led to a greater increase in time to fixate targets in single than dual-target search (Transparent Polygons: single-target:  $F(2,60)=110.53$ ,  $p<.0001$ ,  $ges=.49$ ; dual-target:  $F(2,60)=15.05$ ,  $p<.0001$ ,  $ges=.13$ ; Real-world objects: single-target:  $F(2,60)=80.72$ ,  $p<.0001$ ,  $ges=.29$ ; dual-target:  $F(2,60)=7.67$ ,  $p=.001$ ,  $ges=.06$ ). For Opaque Polygons, there was no significant effect of overlap in dual-target search (single-target:  $F(2,60)=92.81$ ,  $p<.0001$ ,  $ges=.37$ ; dual-target:  $F(2,60)=1.53$ ,  $p=.23$ ,  $ges=.02$ ). In contrast, increasing overlap in the X-ray objects condition increased the time to fixate targets in single-target search, but reduced the time to fixate targets in dual-target search (single-target:  $F(2,60)=3.43$ ,  $p=.038$ ,  $ges=.02$ ; dual-target:  $F(2,60)=3.53$ ,  $p=.035$ ,  $ges=.03$ ).

In summary, it took longer to fixate targets in dual- than single-target search. That said, the tendency for increasing overlap to produce an increase in the time to fixate targets was most apparent in single-target search. Targets were generally fixated quickly in single-target search, but a large amount of overlap slowed down that selection process. Depth appeared to have no influence on the time to fixate targets.

### **Probability of fixating targets.**

The probability of fixating targets also serves as a measure of failures of perceptual selection in search. The initial ANOVA (see Table 4 for ANOVA results and Figure 5 for means) revealed a number of main effects and interactions. Overall, participants were less likely to fixate the target in dual-target than single-target search, and were also less likely to fixate the target as overlap increased. There were four interactions: Depth x Overlap, Stimulus Type x Search Type, Stimulus Type x Overlap and Search Type x Overlap.



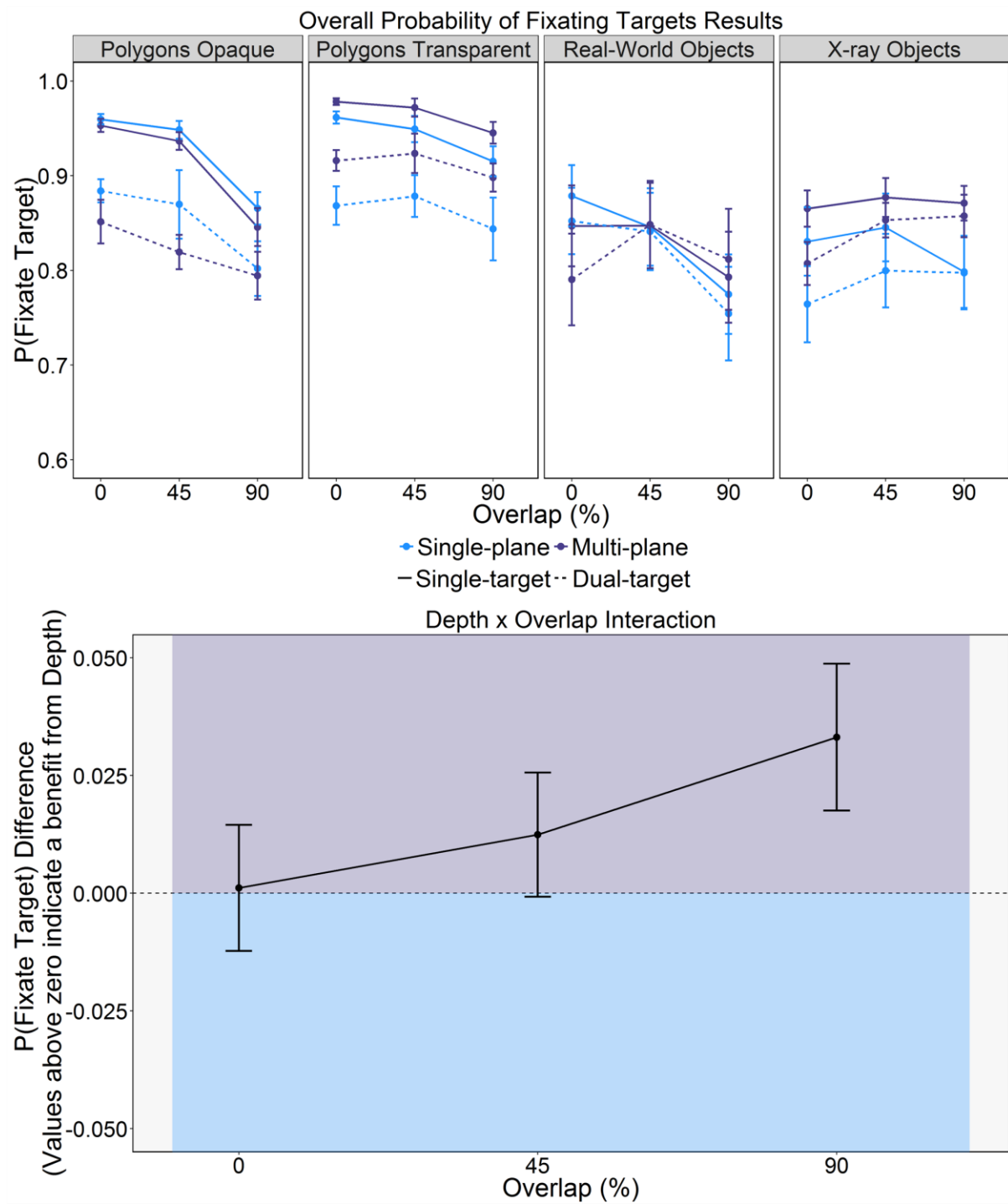


Figure 5. Probability of Fixating Targets (upper panel) and the Probability of Fixating Difference between Multi-plane and Single-plane Search (lower panel) for the different Stimulus Types, levels of Depth and Search Types. Error bars represent SE.

Table 4.

*Main Effects and Interactions for the Initial ANOVA for  $P(\text{Fixate Target})$ .*

Effect	<i>F</i>	<i>df</i>	<i>ges</i>
--------	----------	-----------	------------

Depth	0.66	(1,120)	0.005
Stimulus Type	5.57**	(3,120)	0.10
Search Type	111.93***	(1,120)	0.04
Overlap	42.1***	(2,240)	0.03
Depth x Stimulus Type	0.73	(3,120)	0.01
Depth x Search Type	0.27	(1,120)	0.0001
Stimulus Type x Search Type	10.43***	(3,120)	0.01
Depth x Overlap	4.82**	(2,240)	0.003
Stimulus Type x Overlap	8.61***	(6,240)	0.02
Search Type x Overlap	12.06***	(2,240)	0.004
Depth x Stimulus Type x Search Type	0.91	(3,120)	0.001
Depth x Stimulus Type x Overlap	1.66	(6,240)	0.003
Depth x Search Type x Overlap	0.95	(2,240)	0.0003
Stimulus Type x Search Type x Overlap	1.64	(6,240)	0.002
Depth x Stimulus Type x Search Type x Overlap	1.39	(6,240)	0.001

Note: \*= $p<.05$ , \*\*= $p<.01$ , \*\*\*= $p<.001$

For the Stimulus Type x Search Type and Stimulus Type x Overlap interactions, we examined each Stimulus Type separately. There was evidence of a dual-target cost to the probability of fixating targets in all cases ( $F_s>22$ ,  $p_s<.02$ ) except Real-world stimuli,  $F(1,30)=3.12$ ,  $p=.09$ ,  $ges=.002$ ). The probability of fixating objects reduced between 0% and 90% overlap in all cases other than X-ray objects for which the probability marginally increased ( $F_s>4$ ,  $p_s<.05$ ).

With respect to the Depth x Overlap interaction, the effect of Overlap was attenuated for multi-plane versus single-plane search (multi-plane:  $F(2,120)=13.69$ ,  $p<.001$ ,  $ges=0.015$ ; single-plane:  $F(2,120)=29.48$ ,  $p<.001$ ,  $ges=0.043$ ).

With respect to the Search type x Overlap interaction, the dual-target cost decreased with overlap, though was still significant for all levels of overlap ( $t_s>4$ ,  $p_s<.0001$ ).

In summary, targets were less likely to be fixated in dual- than single-target search, although this difference reduced with overlap and did not hold for X-ray objects. Depth did influence the probability of fixating targets by attenuating the effect of overlap.

### Verification time.

Verification times, defined as the time between first fixating the target and providing a correct response, serves as a measure of perceptual identification during search. There were a number of effects and interactions for this measure (see Table 5 for ANOVA results and

Figure 6 for means). Participants required more time to verify targets in dual-target than single-target search, and as overlap increased. There were two key sets of interactions that subsumed the other effects and interactions: namely, Depth x Overlap and Stimulus Type x Search Type x Overlap.

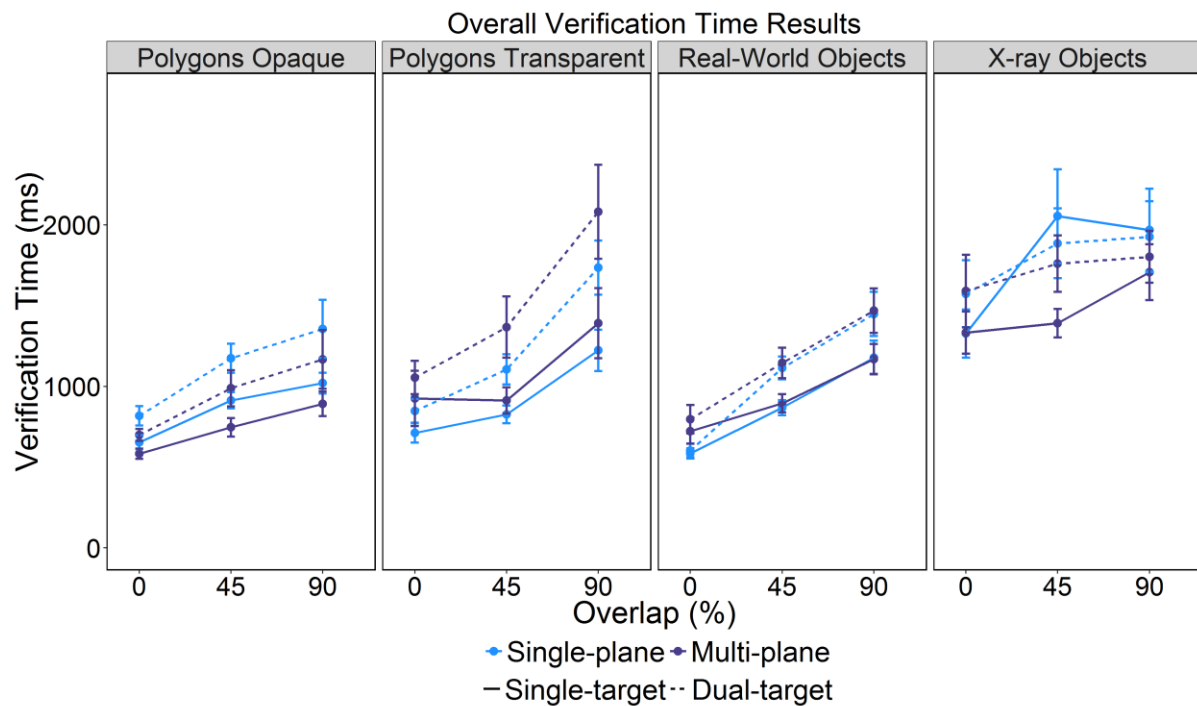


Figure 6. Verification Time for the different Stimulus Types, levels of Depth and Search Types. Error bars represent SE.

Table 5.

Main Effects and Interactions for the Initial ANOVA for Verification Time.

Effect/Interaction	<i>F</i>	<i>df</i>	ges
Depth	0	(1,120)	0.0000
Stimulus Type	15.31* * *	(3,120)	0.21
Search Type	116.6* * *	(1,120)	0.05
Overlap	227.78* * *	(2,240)	0.26
Depth x Stimulus Type	1.49	(3,120)	0.03
Depth x Search Type	1.38	(1,120)	0.001
Stimulus Type x Search Type	7.15* * *	(3,120)	0.01
Depth x Overlap	5.81* *	(2,240)	0.01
Stimulus Type x Overlap	7.53* * *	(6,240)	0.03
Search Type x Overlap	1.48	(2,240)	0.001
Depth x Stimulus Type x Search Type	0.88	(3,120)	0.001

Depth x Stimulus Type x Overlap	0.37	(6,240)	0.002
Depth x Search Type x Overlap	0.9	(2,240)	0.0004
Stimulus Type x Search Type x Overlap	3.83* *	(6,240)	0.01
Depth x Stimulus Type x Search Type x Overlap	1.11	(6,240)	0.002

Note:  $*=p<.05$ ,  $**=p<.01$ ,  $***=p<.001$

For the Stimulus Type x Search Type x Overlap interaction, we examined each of the Stimulus Types separately. The interaction between Search type and Overlap did not reach significance for Opaque and Transparent Polygons ( $F_s<2$ ,  $p_s>.14$ ), though in both, there were main effects of Search Type and Overlap ( $F_s>45$ ,  $p_s<.0001$ ), indicating the verification time was longer in dual-target than single-target search, and that verification time increased as overlap increased. There was an interaction between Search Type and Overlap for the Real-world and X-ray objects ( $F_s>3$ ,  $p_s<.05$ ). This arose because of variations in the dual-target cost for these stimulus types, with no dual-target cost for 0% overlap in the Real-world objects ( $t<1.7$ ), and a dual-target cost for the remaining levels of overlap ( $t_s>7.5$ ,  $p_s<.0001$ ); for the X-ray objects, there was a dual-target cost for 0% overlap only ( $t(31)=2.82$ ,  $p=.008$ ).

For the Depth x Overlap interaction, while the interaction was significant, no pairwise contrasts reached significance ( $t_s<1.2$ ,  $p_s>.6$ ).

In summary, verification times were longer in dual- than single-target search and increased with Overlap. However, the difference in verification times between single- and dual-target search reduced with Overlap in the Opaque Polygons and X-ray objects conditions.

### **Probability of identifying targets after fixating them.**

The probability of identifying targets after fixating them also serves as a measure of failures of perceptual identification during search. The initial ANOVA, as with the previous measures, revealed a number of effects and interactions (see Table 6 for ANOVA results and Figure 7 for means). Participants were more likely to identify fixated targets in single-target than dual-target search, and were less likely to identify fixated targets as overlap increased. There were a number of interactions that were subsumed into two other interactions: namely, a Depth x Overlap interaction and a Stimulus Type x Search Type x Overlap interaction.

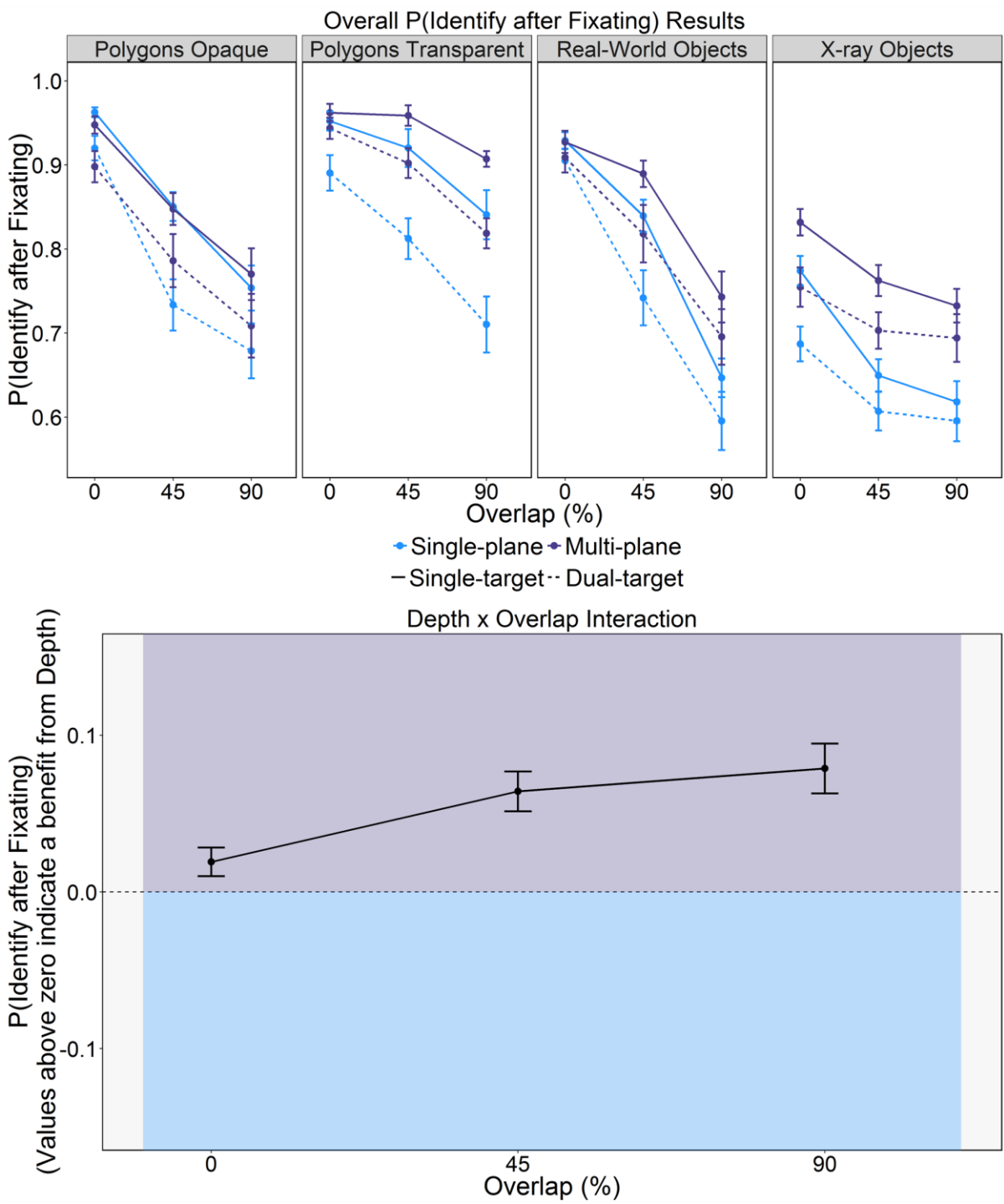


Figure 7. Probability of Identifying Targets after Fixating them (upper panel) and the Mean Probability of Identifying Targets after Fixating them Difference between Multi-plane and Single-plane Search (lower panel), for the different Stimulus Types, levels of Depth and Search Types. Error bars represent SE.

Table 6.

*Significant Main Effects and Interactions for the Initial ANOVA for the Probability of Fixating and Identifying Targets.*

Effect/Interaction	<i>F</i>	<i>df</i>	<i>ges</i>
Depth	20.13***	(1,120)	0.08
Stimulus Type	40.12***	(3,120)	0.36
Search Type	167.93***	(1,120)	0.11
Overlap	302.89***	(2,240)	0.37
Depth x Stimulus Type	1.94	(3,120)	0.03
Depth x Search Type	3.29	(1,120)	0.002
Stimulus Type x Search Type	1.51	(3,120)	0.003
Depth x Overlap	10.37***	(2,240)	0.02
Stimulus Type x Overlap	16.3***	(6,240)	0.09
Search Type x Overlap	4.49*	(2,240)	0.005
Depth x Stimulus Type x Search Type	1.28	(3,120)	0.003
Depth x Stimulus Type x Overlap	0.5	(6,240)	0.003
Depth x Search Type x Overlap	0.51	(2,240)	0.001
Stimulus Type x Search Type x Overlap	4.31***	(6,240)	0.01
Depth x Stimulus Type x Search Type x Overlap	0.45	(6,240)	0.001

*Note:* \*= $p < .05$ , \*\*= $p < .01$ , \*\*\*= $p < .001$

We broke down the Stimulus Type x Search Type x Overlap interaction by looking at each Stimulus type separately, in line with our previous analyses. The interaction between Search type and Overlap was significant for Transparent Polygons, Real-world objects, and X-ray objects ( $F_s > 3.5$ ,  $p_s < .05$ ) but not for Opaque Polygons ( $F(2,60) = 2.82$ ,  $p = .07$ ,  $ges = .008$ ). For Transparent Polygons, overlap had a reduced effect on the probability of identifying targets after fixating them in dual- than single-target search (single-target:  $F(2,60) = 36.57$ ,  $p < .0001$ ,  $ges = .35$ ; dual-target:  $F(2,60) = 27.15$ ,  $p < .0001$ ,  $ges = .22$ ), whereas for the remaining stimulus types, overlap had a reduced effect in single- than dual-target search (Real-world single-target:  $F(2,60) = 213.03$ ,  $p < .0001$ ,  $ges = .64$ ; Real-world dual-target:  $F(2,60) = 62.02$ ,  $p < .0001$ ,  $ges = .48$ ; X-ray single-target:  $F(2,60) = 54.54$ ,  $p < .0001$ ,  $ges = .34$ ; X-ray dual-target:  $F(2,60) = 13.64$ ,  $p < .0001$ ,  $ges = .12$ ).

For the Depth x Overlap interaction, we examined each Overlap level separately using *t*-tests. These revealed that participants were more likely to identify targets after fixating them in multi-plane than single-plane displays in the 45% and 90% Overlap conditions only ( $t_s > 3.2$ ,  $p_s < .01$ ). Furthermore, in line with the probability of fixating targets, the effect size

for Overlap was attenuated for multi-plane compared with single-plane displays (multi-plane:  $F(2,120)=110.29$ ,  $p<.001$ ,  $g=0.29$ ; single-plane:  $F(2,120)=196.9$ ,  $p<.001$ ,  $g=0.44$ ).

In summary, the probability of identifying targets after fixating them was higher in single- than dual-target search and reduced with overlap. The extent of the reduction with Overlap varied across single- and dual-target search and stimulus conditions. Importantly, as overlap increased, the probability of identifying fixated targets was higher when objects appeared in depth.

## Summary of Results

We will now briefly summarize the results, focusing on the effects of overlap, single/dual-target search, and the effects of adding depth to the displays.

Increasing overlap impaired search performance across all measures. Increases in overlap resulted in a decrease in response accuracy, coupled with an increase in response times. As overlap increased, participants were slower and less likely to fixate targets, demonstrating that overlap impaired perceptual selection processes. As overlap increased, verification time increased, and the probability of identifying targets after fixating them decreased, demonstrating that overlap impaired perceptual identification processes.

Turning to the comparisons of single- and dual-target cost, throughout we found evidence of a dual-target cost in terms of dual-target search having longer RTs and lower response accuracy than single-target search, in line with previous research. As with the effects of overlap, compared to single-target search, dual-target search involved an increase in the time taken to fixate targets, coupled with a reduction in the probability of fixating targets, demonstrating that dual-target search impairs perceptual selection processes. Moreover, when compared to single-target search, dual-target search resulted in an increase in verification time, as well as a reduction in the probability of identifying targets after fixating them, demonstrating that dual-target search impairs perceptual identification processes.

Finally, with regards to the effects of adding depth to the displays, we found that adding depth to the displays improved response accuracy for the transparent stimuli (transparent polygons, X-ray images), and for target-present real-world stimuli, but had little effect on RTs for all stimulus types. When depth did influence eye movement behavior, it had a blanket effect across all stimulus types, rather than being restricted to transparent displays only. We found that adding depth to the displays did not influence the time to fixate targets, but did attenuate the effects of overlap for the probability of fixating targets. This suggests to

some extent that depth aids perceptual selection processes. Adding depth to the displays did not influence verification times, though it did increase the probability that participants would identify targets after fixating them.

### Discussion

In the present experiments, we examined two aspects of 'everyday' visual searches that have remained largely ignored by current models and theories of search: namely, depth, and overlap. Although real-world search tasks are replete with variations in depth and overlap amongst objects, they are the exception, rather than the norm, in laboratory-based search experiments. We expected that adding depth to the displays would help participants to better segregate overlapping objects and identify them, and, as a result, improve search performance. We engaged participants in visual searches wherein overlap was varied (0%, 45%, and 90%), and asked them to search for one target (single-target search) or two targets (dual-target search). Half of the participants searched displays where the objects were presented on different depth planes to one another (the multi-plane condition); the remaining participants searched displays where the objects were presented on a single depth plane (the single-plane condition). To ensure that our results would generalise across stimulus types, participants searched opaque displays (opaque polygons, real-world objects) and transparent displays (transparent polygons, X-ray objects). We examined behavioural measures (response accuracy, response times), as well as eye movement measures focusing on failures of perceptual selection (time to fixate targets, probability of fixating targets), combined with failures of perceptual identification (verification time, probability of identifying targets after fixating them). We predicted that overlap would increase RTs and decrease response accuracy, and that it would also increase errors of perceptual selection and identification. We also predicted that adding depth to the displays would improve behavioural performance and reduce errors of perceptual selection and identification, though we expected that these improvements would be focused, or even restricted, to the transparent displays. This, we predicted, would occur because regions where transparent stimuli overlap introduce ambiguity as to object identity, whilst regions where opaque stimuli overlap entirely remove information relating to object identity (aside from for the object that is 'above' the other objects). We will now discuss the results in relation to overlap, dual-target search, and depth.

As expected, overlap did impair performance by increasing RTs and decreasing response accuracy, in line with previous research. Given that previous studies of overlap during visual search (Bravo & Farid, 2004a, 2004b, 2006; Hillstrom et al., 2013) have not



examined errors of perceptual selection and identification, our analyses in this regard tell us something new about the precise manner in which overlap impairs performance. Since overlap is commonplace during everyday visual searches, one might have expected search to be relatively robust against the effects of overlap. Instead, our analyses of participants' eye movement behaviour make a number of novel insights into how increasing overlap impairs search performance. We found that participants were less likely and slower to fixate targets as overlap increased, demonstrating that overlap does indeed impair perceptual selection processes in visual search. Perceptual selection is likely impaired because overlap can either remove (in the case of opaque objects) or obscure (in the case of transparent objects) features that are required to make a veridical identification of the target object (e.g., see Gosselin & Schyns, 2001; Schyns, 1998). These findings were in line with our predictions, as was the finding that the effects of overlap were more damaging for opaque than transparent objects, since in transparent overlapping objects, the diagnostic features are still visible but need careful interpretation for successful selection. Similar findings emerged for the measures of perceptual identification: overlap increased verification times and also reduced the likelihood that participants would identify fixated targets. These findings go beyond previous studies of overlap by identifying the specific effects on perceptual selection and identification that underlie shifts in performance.

While not the issue of direct interest, we feel it important to summarise the results from dual-target search for completeness. We found evidence of a dual-target cost in almost all cases. Dual-target search was slower and had reduced response accuracy compared to single-target search for all stimulus types, and for both target-present and target-absent trials. The only exception was for response accuracy in the 90% overlap X-ray stimuli, for which there was no dual-target cost. This was most likely a floor effect since response accuracy here was very low indeed, even for single-target search. Moreover, a dual-target cost emerged across all stimulus types in all of the eye movement measures, demonstrating that, in line with previous research (Menneer et al., 2004, 2007), dual-target search results in an increase in errors of perceptual selection and perceptual identification.

The most important and novel results from the present set of experiments focus on how the presence of depth in overlapping displays modulates visual search behaviour and performance. We expected that the presence of depth would aid search by facilitating object segmentation processes. We found that the presence of depth in the displays improved response accuracy for the transparent display types (Transparent Polygons, X-ray images), as well as for the target-present trials in the Real-world stimuli. For RTs, the pattern of results

was less clear-cut: the presence of depth ameliorated the effects of overlap to some extent, but had no other effects. Although the effects of adding depth to the displays focused primarily on the transparent stimulus types for response accuracy, the effects of adding depth upon the eye movement measures had similar effects across all stimulus types. In terms of perceptual selection processes, we found that the presence of depth had no influence upon the time taken to fixate targets, although the presence of depth did reduce the effect of overlap on the probability of fixating targets. In terms of perceptual identification processes, the presence of depth in the displays reduced the effect size of overlap for verification times, and increased the probability of identifying targets after fixating them for 45% and 90% overlap. Overall, we therefore found evidence that the presence of depth aided response accuracy for some stimulus types, with some evidence that the presence of depth aided perceptual selection (for the probability of fixating targets, not the time to fixate targets), and stronger evidence that the presence of depth aided perceptual identification (verification time, probability of identifying targets after fixating them), though this was true across all stimulus types. Improving perceptual selection and perceptual identification processes can both explain why response accuracy increased when depth was added to the displays, though the influence that depth had upon eye movement behaviour only translated to a significant benefit in search performance for some of the stimulus types. Indeed, the fact that the benefits to search performance only emerged for some stimulus types, and mostly when the displays contained overlapping objects, helps to explain why previous studies of depth have shown such mixed results (McIntire, Havig, & Geiselman, 2014): the presence of depth is only beneficial under a certain set of conditions, and even in some cases when depth influences eye movement behaviour, that does not necessarily lead to a shift in behavioural performance.

From a theoretical perspective, our experiments are the first to show that presentation of objects in depth can aid performance when objects overlap each other and are transparent, and primarily improve object identification processes. Current models of search tend to be based upon simple tasks wherein participants search for simple targets that are easily segmented from other objects, and have fully visible features. As a consequence, object identification has not been captured in detail within the tasks that these models are based on, and the search models themselves generally do not consider the ways in which object identification can fail; rather they focus on guidance of attention towards fully visible targets (Thornton & Gilden, 2007; Wolfe, 2007). However, the present set of studies adds to a growing body of research that explores failures of target identification in search, which can

occur even after successful guidance to the target. Models along these lines incorporate a probabilistic mechanism by which object identification can fail (Godwin, Reichle, et al., in press; Wolfe & Van Wert, 2010), but the present set of studies highlight the need for a better understanding of exactly how and why object identification processing can fail in search. We suggest that combining existing models with the use of overlapping stimuli could provide a rich source of new information regarding the moment-to-moment processing of object identification in the context of visual search tasks.

From a practical perspective, these findings are important because they help to inform some of the complex challenges facing real world searches (i.e., overlapping objects, the dual-target cost and transparency), but also offer a positive outlook in the sense that adding depth information to displays containing overlapping objects could improve performance. These findings from a broad and controlled set of conditions is particularly important given the mixed evidence in favour of using depth information in displays (McIntire et al., 2014), and given the proliferation of technology that is available to present depth information on displays.

Since our findings relating to the presence of depth in overlapping displays are new in the sense that previous research has not examined perceptual selection and identification processes in visual search along these lines, it is important that future studies replicate these findings. This is particularly the case given the mixed evidence relating to the benefits of adding depth to displays that has already been highlighted in recent reviews. Moreover, although we counterbalanced the various factors and conditions within our experiments, it still remains possible that the effects that we have observed occur due to the fact that participants were inexperienced with regards to searching overlapping displays containing depth information. In fact, given that everyday visual searches tend to involve overlapping displays containing depth information, it could even be argued that participants may have already been ‘experts’ in terms of the multi-plane condition, and the effects that we observed emerged as a result of participants’ inexperience with searching single-plane overlapping displays. Therefore, to determine whether the effects of adding depth to displays are fragile (or short-lived) or long-lasting, we plan in future research to give participants extensive experience with searching through single-plane and multi-plane displays. This would also serve as a replication of the present set of experiments which found that the presence of depth in visual search displays can, in some situations, improve visual search performance.

## References

- Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, 37(3), 379–84. doi:10.3758/BF03192707
- Barrett, D. J. K., & Zobay, O. (2014). Attentional Control via Parallel Target-Templates in Dual-Target Search. *PloS One*, 9(1), e86848. doi:10.1371/journal.pone.0086848
- Beck, V. M., Hollingworth, A., & Luck, S. J. (2012). Simultaneous control of attention by multiple working memory representations. *Psychological Science*, 23(8), 887–98. doi:10.1177/0956797612439068
- Becker, S. I. (2011). Determinants of dwell time in visual search: similarity of perceptual difficulty? *PLoS One*, 6(3), 1–5. doi:10.1371/journal.pone.0017740
- Bravo, M. J., & Farid, H. (2004a). Recognizing and segmenting objects in clutter. *Vision Research*, 44(4), 385–96. doi:10.1016/j.visres.2003.09.031
- Bravo, M. J., & Farid, H. (2004b). Search for a category target in clutter. *Perception*, 33(6), 643–52.
- Bravo, M. J., & Farid, H. (2006). Object recognition in dense clutter. *Perception & Psychophysics*, 68(6), 911–8. doi:10.3758/BF03193354
- Cain, M. S., Adamo, S. H., & Mitroff, S. R. (2013). A taxonomy of errors in multiple-target visual search. *Visual Cognition*, 21(7), 899–921. doi:10.1080/13506285.2013.843627
- Castelhano, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an unspecified target, but only in identification, and not in attentional guidance. *Psychonomic Bulletin & Review*, 15, 795–801. doi:10.3758/PBR.15.4.795
- Eckstein, M. P. (2011). Visual search : A retrospective. *Journal of Vision*, 11(5:14), 1–36. doi:10.1167/11.5.14.
- Finlayson, N. J., Remington, R. W., Retell, J. D., & Grove, P. M. (2013). Segmentation by depth does not always facilitate visual search. *Journal of Vision*, 13(8), 1–14. doi:10.1167/13.8.11
- Godwin, H. J. (2013). Functions for organising and analysing eye-tracking data.
- Godwin, H. J., Holliman, N. S., Menneer, T., Liversedge, S. P., Cave, K. R., & Donnelly, N. (in press). Assessing the benefits of stereoscopic displays to visual search: methodology and initial findings. In *Stereoscopic Displays and Applications XVI*. San Francisco, United States.
- Godwin, H. J., Hout, M. C., & Menneer, T. (2014). Visual similarity is stronger than semantic similarity in guiding visual search for numbers. *Psychonomic Bulletin and Review*, 21(3), 689–95. doi:10.3758/s13423-013-0547-4
- Godwin, H. J., Menneer, T., Cave, K. R., & Donnelly, N. (2010). Dual-target search for high and low prevalence X-ray threat targets. *Visual Cognition*, 18(10), 1439–1493. doi:doi:10.1080/13506285.2010.500605
- Godwin, H. J., Menneer, T., Riggs, C. A., Cave, K. R., & Donnelly, N. (2015). Perceptual Failures in the Selection and Identification of Low-Prevalence Targets in Relative Prevalence Visual Search. *Attention, Perception & Psychophysics*, 77(1), 150–9. doi:10.3758/s13414-014-0762-8
- Godwin, H. J., Menneer, T., Riggs, C. A., Taunton, D., Cave, K. R., & Donnelly, N. (2015). Understanding the contribution of target repetition and target expectation to the

- emergence of the prevalence effect in visual search. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-015-0970-9
- Godwin, H. J., Reichle, E. D., & Menneer, T. (in press). Modeling Lag-2 Revisits to Understand Trade-Offs in Mixed Control of Fixation Termination During Visual Search. *Cognitive Science*.
- Godwin, H. J., Walenchok, S. C., Houpt, J. W., Hout, M. C., & Goldinger, S. D. (2015). Faster Than the Speed of Rejection : Object Identification Processes During Visual Search for Multiple Targets. *Journal of Experimental Psychology : Human Perception and Performance*.
- Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research*, 41(17), 2261–71. doi:10.1016/S0042-6989(01)00097-9
- Grubert, A., & Eimer, M. (2016). All Set, Indeed! N2pc Components Reveal Simultaneous Attentional Control Settings for Multiple Target Colors. *Journal of Experimental Psychology: Human Perception and Performance*. doi:10.1037/xhp0000221
- He, Z., & Nakayama, K. (1992). Surfaces versus features in visual search. *Nature*, 359, 231–233. doi:10.1038/359231a0
- Hillstrom, A. P., Wakefield, H., & Scholey, H. (2013). The effect of transparency on recognition of overlapping objects. *Journal of Experimental Psychology. Applied*, 19(2), 158–70. doi:10.1037/a0033367
- Hout, M. C., Walenchok, S. C., Goldinger, S. D., & Wolfe, J. M. (2015). Failures of perception in the low-prevalence effect: Evidence from active and passive visual search. *Journal of Experimental Psychology: Human Perception and Performance*. doi:10.1037/xhp0000053
- Houtkamp, R., & Roelfsema, P. R. (2009). Matching of visual input to only one item at any one time. *Psychological Research*, 73(3), 317–26. doi:10.1007/s00426-008-0157-3
- Irons, J. L., Folk, C. L., & Remington, R. W. (2012). All set! Evidence of simultaneous attentional control settings for multiple target colors. *Journal of Experimental Psychology. Human Perception and Performance*, 38(3), 758–75. doi:10.1037/a0026578
- Ishihara, S. (1964). *Tests for colour-blindness (24 plates)*. Tokyo, Japan: Kanehara Shuppan Co., Ltd.
- Lawrence, M. A. (2015). Easy Analysis and Visualization of Factorial Experiments.
- Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, 4(1), 6–14. doi:10.1016/S1364-6613(99)01418-7
- Luria, S. M., & Strauss, M. S. (1975). Eye movements during search for coded and uncoded targets. *Perception and Psychophysics*, 17(3), 303–308. doi:10.3758/BF03203215
- McIntire, J. P., Havig, P. R., & Geiselman, E. E. (2012). What is 3D good for? A review of human performance on stereoscopic 3D displays. In *Head- and Helmet-Mounted Displays XVII; and Display Technologies and Applications for Defense, Security, and Avionics VI*, 83830X. SPIE. doi:10.1117/12.920017
- McIntire, J. P., Havig, P. R., & Geiselman, E. E. (2014). Stereoscopic 3D Displays and Human Performance: A Comprehensive Review. *Displays*, 35(1), 18–26. doi:10.1016/j.displa.2013.10.004

- McSorley, E., & Findlay, J. M. (2001). Visual search in depth. *Vision Research*, 41(25-26), 3487–96. doi:10.1016/S0042-6989(01)00197-3
- Menneer, T., Barrett, D. J. K., Phillips, L., Donnelly, N., & Cave, K. R. (2004). Search efficiency for multiple targets. *Cognitive Technology*, 9, 22–25.
- Menneer, T., Barrett, D. J. K., Phillips, L., Donnelly, N., & Cave, K. R. (2007). Costs in searching for two targets: Dividing search across target types could improve airport security screening. *Applied Cognitive Psychology*, 21(7), 915–932. doi:10.1002/acp.1305
- Menneer, T., Cave, K. R., & Donnelly, N. (2009). The cost of search for multiple targets: the effects of practice and target similarity. *Journal of Experimental Psychology: Applied*, 15, 125–139. doi:10.1037/a0015331
- Menneer, T., Stroud, M. J., Cave, K. R., Li, X., Godwin, H. J., Liversedge, S. P., & Donnelly, N. (2012). Search for two categories of target produces fewer fixations to target-color items. *Journal of Experimental Psychology: Applied*, 18(4), 404–18. doi:10.1037/a0031032
- Moore, C. M., & Osman, a M. (1993). Looking for two targets at the same time: one search or two? *Perception & Psychophysics*, 53(4), 381–90. doi:10.3758/BF03206781
- Nakayama, K., Shimojo, S., & Silverman, G. (1989). Stereoscopic depth: its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception*, 18(1), 55–68. doi:10.1068/p180055
- Nakayama, K., & Silverman, G. H. (1986). Serial and Parallel Processing of Visual Feature Conjunctions. *Nature*, 320(6059), 264–265. doi:10.1038/320264a0
- Nodine, C. F., & Kundel, L. (1987). Using eye movements to study visual search and to improve tumor detection. *RadioGraphics*, 7(6), 1241–1250.
- O’Toole, A. J., & Walker, C. L. (1997). On the preattentive accessibility of stereoscopic disparity: evidence from visual search. *Perception & Psychophysics*, 59(2), 202–18. doi:10.3758/BF03211889
- R Development Core Team. (2013). R: A language and environment for statistical computing. (R. D. C. Team, Ed.) *R Foundation for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506. doi:10.1080/17470210902816461
- Rayner, K., Smith, T. J., Malcolm, G. L., & Henderson, J. M. (2009). Eye movements and visual encoding during scene perception. *Psychological Science*, 20(1), 6–10. doi:10.1111/j.1467-9280.2008.02243.x
- Schwark, J. D., MacDonald, J., Sandry, J., & Dolgov, I. (2013). Prevalence-based decisions undermine visual search. *Visual Cognition*, 21(5), 541–568. doi:10.1080/13506285.2013.811135
- Schyns, P. G. (1998). Diagnostic recognition: task constraints, object information, and their interactions. *Cognition*, 67(1-2), 147–79. doi:10.1016/S0010-0277(98)00016-X
- Sobel, K. V., Puri, A. M., & Hogan, J. (2014). Target grouping in visual search for multiple digits. *Attention, Perception, & Psychophysics*, 77(1), 67–77. doi:10.3758/s13414-014-0761-9

- Stroud, M. J., Menneer, T., Cave, K. R., & Donnelly, N. (2012). Using the dual-target cost to explore the nature of search target representations. *Journal of Experimental Psychology: Human Perception and Performance*, 38(1), 113–122. doi:10.1037/a0025887
- Stroud, M. J., Menneer, T., Cave, K. R., Donnelly, N., & Rayner, K. (2011). Search for multiple targets of different colours: Misguided eye movements reveal a reduction of colour selectivity. *Applied Cognitive Psychology*, 25(6), 971–982. doi:10.1002/acp.1790
- Thornton, T. L., & Gilden, D. L. (2007). Parallel and serial processes in visual search. *Psychological Review*, 114(1), 71–103. doi:10.1037/0033-295X.114.1.71
- van Beurden, M. H. P. H., van Hoey, G., Hatzakis, H., & Ijsselstein, W. A. (2009). Stereoscopic displays in medical domains: a review of perception and performance effects. In *Proceedings SPIE 7240, Human Vision and Electronic Imaging XIV, 72400A* (Vol. 7240). doi:10.1117/12.817748
- Wolfe, J. M. (2007). Guided Search 4.0: Current Progress with a model of visual search. In W. Gray (Ed.), *Integrated Models of Cognitive Systems* (pp. 99–119). New York: Oxford.
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3), 419–433. doi:10.1037/0096-1523.15.3.419
- Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology*, 20(2), 121–4. doi:10.1016/j.cub.2009.11.066
- Zelinsky, G. (2008). A theory of eye movements during target acquisition. *Psychological Review*, 115(4), 787–835. doi:10.1037/a0013118