RUNNING HEAD: TRANING WITH DEPTH IN VISUAL SEARCH

# Experience with Searching in Displays containing Depth Improves Search Performance by Training Participants to Search more Exhaustively

Hayward J. Godwin[a], Tamaryn Menneer[b], Simon P. Liversedge[c], Kyle R. Cave[d], Nick S. Holliman[e] & Nick Donnelly[f]

[a]University of Southampton

[b]University of Exeter

[c]University of Central Lancashire

[d]University of Massachusetts Amherst

[e]Newcastle University

[f]Liverpool Hope University

Author Note

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

Copyright is held by Elsevier. This prepublication version may not exactly replicate the final version published in Acta Psychologica, October 2020, and is not the copy of record.

Abstract

In a typical visual search task, participants search for single targets amongst displays containing non-overlapping objects that are presented on a single depth plane. Recent work has begun to examine displays containing overlapping objects that are presented on different depth planes to one another. It has been found that searching displays containing depth improves response accuracy by making participants more likely to fixate targets and to identify targets after fixating them. Here we extended this previous research by seeking first of all to replicate the previous pattern of results, and then to determine whether extensive training using depth in search transfers to two-dimensional displays. We provided participants with sixteen sessions of training with displays containing transparent overlapping objects presented in depth, and found a similar pattern of results to our previous study. We also found evidence that some performance improvements from the depth training transferred to search of two-dimensional displays that did not contain depth. Further examinations revealed that participants learn to search more exhaustively (i.e., search for longer) in displays containing depth. We conclude that depth does influence search performance but the influences depend very much on the stimuli and the degree of overlap within them.

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

## 1.0 Experience with Searching in Displays containing Depth Improves Search Performance by Training Participants to Search more Exhaustively

Human observers are often engaged in visual search tasks that are a far cry from the visual environments that our visual systems evolved to process. The complex visualisations made available by modern technologies require searchers to learn how to interpret images that do not necessarily follow the same 'rules' that the real world follows (Muhl-Richardson et al., 2018; Richardson et al., 2018). For example, in airport baggage screening, screening personnel are required to examine X-ray images which contain a wide array of transparent overlapping objects (for a review, see Donnelly, Muhl-richardson, Godwin, & Cave, 2019). The fact that X-ray images are transparent and overlapping sets them apart from 'everyday' searches we routinely conduct-–such as searching for a set of keys on a messy desk or for the face of a friend in a crowd--wherein objects are typically opaque and overlapping (Hillstrom, Wakefield, & Scholey, 2013). Despite their complexity, human searchers are able to learn how to search these images and interpret them (Schwaninger et al., 2008).

Learning how to interpret these images not only requires an understanding of the appearance of threat/prohibited items from different orientations, but also how to disentangle regions where multiple transparent objects overlap. Typically, airport screeners learn how to search these images using two-dimensional displays, and then conduct their searches using two-dimensional displays (see also Buser, Sterchi, & Schwaninger, 2020; Sterchi, Hättenschwiler, & Schwaninger, 2019). Here, we addressed whether there would be benefits to search performance by training searchers to search displays containing transparent overlapping objects wherein the objects are presented upon different depth planes to one another. As a stronger test of this possibility, we also tested whether the benefits from training to search in depth transferred across to search two-dimensional displays as well. The core idea behind this is the fact that, during training, if overlapping objects are presented on different depth planes to one another, then it is possible that depth can be used as a cue to learn how to segment and segregate those objects, thereby facilitating search.

At a theoretical level, our goal was to push forward what is known regarding training and experience searching displays containing depth. More than that, however, is the question of whether training in displays containing depth aids in the segmentation and identification of overlapping objects. On this point, the current literature is relatively unclear. Indeed, there have been considerable efforts geared towards addressing whether adding depth to search displays can aid performance. Within the context of standard search tasks, the role that the presence of depth plays in influencing search performance has been primarily studied with target objects defined as being at a particular depth (Finlayson, Remington, Retell, & Grove, 2013; He & Nakayama, 1992; McSorley & Findlay, 2001; Nakayama, Shimojo, & Silverman, 1989; Nakayama & Silverman, 1986; O'Toole & Walker, 1997) , as opposed to training participants to search for targets that could appear at any of a number of depths. Elsewhere, studies that have examined the potential benefits of adding depth to displays during training have found mixed evidence to date. Recent reviews covering the medical domain (van Beurden, IJsselsteijn, & Juola, 2012), and other domains (McIntire, Havig, & Geiselman, 2014) have found some evidence in favour of using depth to aid training. However, the benefits of depth for visual search are still uncertain, both in medical training and in other domains, such as driving vehicles (e.g., see Szczerba & Hersberger, 2014).

Overall, then, there is a clear need for more studies of the influence of training in adding depth on visual search, and exactly what and how those benefits are conferred in terms of the moment-to-moment information processing that takes place when examining displays containing transparent overlapping objects. The present study serves as an extension of a recent and detailed series of studies which examined how the presence of depth in search displays influences performance (Godwin et al., 2017). Our goal was to examine whether or not distributing items across different depth planes aided visual search performance across a range of different stimulus types, and using varying levels of overlap. To ensure that our results would generalize across different search tasks and domains, we used four stimulus types: opaque polygons, transparent polygons, real-world household items, and transparent images from airport X-ray baggage screening. We predicted that the presence of depth would be primarily of benefit when objects were overlapping and transparent. This prediction was based upon the notion that participants could use the depth information as a cue to segregate and more readily identify overlapping objects. However, we expected that such facilitation would primarily appear with transparent objects, for which there would be sufficient visual information available to allow this segregation to take place.

In order to address whether the presence of depth in displays influenced visual search behaviour, we analysed the data from our previous studies in two ways, focusing on behavioural performance (response accuracy, RTs), as well as eye movement metrics. We found that the presence of depth in the displays had little effect upon RTs. However, the presence of depth did improve response accuracy during search, for the transparent stimulus types (transparent polygons, X-ray objects), and also for the real-world stimuli when a target was present. Moreover, we examined the eye movements of participants as they searched, examining failures of *perceptual selection* and *perceptual identification*. These measures have become increasingly popular as a method for pinpointing how and why participants fail to detect targets (Cain, Adamo, & Mitroff, 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), particularly in complex search tasks (Cain et al., 2013; Nodine & Kundel, 1987; Schwark, MacDonald, Sandry, & Dolgov, 2013). Perceptual selection (measured using the time to fixate targets and the probability of fixating targets) is important because if participants fail to fixate targets, or are slow to fixate targets, then those targets are likely to be missed. Perceptual identification (measured using verification time, defined as the time between fixating the target and responding, as also measured by the probability of identifying targets after fixating them) is important because if participants fail to identify targets, or are slow to identify targets, then targets are again likely to be missed.

Although the presence of depth in our previous experiments influenced response accuracy only for some stimulus types in our previous study, including transparent polygons and X-ray images, the presence of depth had a blanket effect upon eye movements across the range of metrics. The presence of depth did not influence the time to fixate targets, but did attenuate the effects of overlap on the probability of fixating targets, suggesting that depth had some effect upon perceptual selection processes. In addition, the presence of depth in the displays did not influence verification times, but did raise the probability that participants would identify targets after having fixated them, suggesting that depth had some effect upon perceptual identification processes. Clearly, then, the presence of depth in the displays does influence visual search, particularly eye movements, but this only translates to a shift in response accuracy under certain conditions.

We know from our previous studies that the presence of depth can indeed aid search performance in the short term, but it is also important to understand the longer-term effects of depth in relation to aiding search performance. At a general level, depth might benefit search performance in two ways: depth information might improve performance by accelerating improvements that would be gained over time, or it might provide a permanent, long-term benefit to search performance, potentially by training searchers to adopt new strategies that they would never have learned without the presence of depth information in the displays.

To better understand these issues, we therefore engaged participants in a series of 17 sessions searching images from X-ray baggage screening that we have used in previous studies (see Table 1). Objects were arranged into overlapping 'clusters' and could overlap up to 90 % of their visual area with other objects. Half of the participants were trained using objects presented on multiple planes of depth (the *multi-plane* condition); the remaining half were trained using objects presented on a single plane of depth (the *single-plane* condition), in a similar approach to our previous study (Godwin et al., 2017). Every four sessions, participants' eye movements were recorded as they searched. In the final session, during which eye movements were again recorded, all participants were presented with displays in 2D in order to directly test for transfer effects from single- or multi-plane to standard, 2D displays.

Table 1.

*Session Plan for the Study.*

| Session | Depth | Includes Eye-Tracking? | Part of Training Effects ANOVA? | Part of Transfer Effects ANOVA? |
|---|---|---|---|---|
| 1 | Single/multi | | Yes | |
| 2 | Single/multi | | Yes | |
| 3 | Single/multi | | Yes | |
| 4 | Single/multi | Yes | Yes | |

| | | | | |
|---|---|---|---|---|
| 5 | Single/multi | | Yes | |
| 6 | Single/multi | | Yes | |
| 7 | Single/multi | | Yes | |
| 8 | Single/multi | Yes | Yes | |
| 9 | Single/multi | | Yes | |
| 10 | Single/multi | | Yes | |
| 11 | Single/multi | | Yes | |
| 12 | Single/multi | Yes | Yes | |
| 13 | Single/multi | | Yes | |
| 14 | Single/multi | | Yes | |
| 15 | Single/multi | | Yes | |
| 16 | Single/multi | Yes | Yes | Yes |
| 17 | 2D (Transfer) | Yes | | Yes |

Following our previous study, we engaged participants in both single-target and dual-target search. Searching for two targets at once results in a *dual-target cost*. The dual-target cost manifests as a reduction in response accuracy, coupled with an increase in response time (Menneer, Barrett, Phillips, Donnelly, & Cave, 2007). The cost is not eliminated with extensive training/practice (Menneer, Cave, & Donnelly, 2009; Menneer et al., 2012), and is a consequence of impairments in search guidance when selecting target-similar objects (Barrett & Zobay, 2014; Stroud, Menneer, Cave, & Donnelly, 2012), as well as fundamental limitations in determining whether attended objects in a display are one of two potential target items (Godwin, Walenchok, Houpt, Hout, & Goldinger, 2015). Here, we also asked participants to engage in dual-target search to determine whether training in depth, and the expected subsequent improvement in search performance, could ameliorate the dual-target cost.

To summarise, our first objective was to replicate the findings of our previous study (Godwin et al., 2017). To reiterate, we previously found no effects from the presence of depth upon RTs, though the presence of depth did improve response accuracy. We expected a similar result from the present study. In terms of eye-movements, we also expected, as with our previous study, that participants would be more likely to fixate targets in the presence of depth, and also show evidence of being more likely to identify targets after fixating them in the presence of depth. Our second objective that built upon our previous study was to examine training effects that arise from searching displays containing the presence of depth, and to determine whether those effects are long-lasting (i.e., they transfer over to searching novel, two-dimensional displays). At a theoretical level, this is an important issue for understanding the mechanics of how and why depth influences visual search. At a practical level, it is important to show whether training using depth information might improve performance in day-to-day searches that do not include depth.

# 2.0 Method

## 2.1 Participants

Prior to taking part in the study, participants completed a series of screening tests. These included tests for color vision (Ishihara, 1964) , as well as 3D depth perception (a score of nine for the Wirt Circles component of the Titmus Stereo test). Participants were undergraduates, postgraduates and staff who took part for payment or course credit. A total of 44 participants took part in the study. Ethical approval was obtained by the University of Southampton Psychology Ethics Committee prior to commencing the study, and informed consent was obtained from all participants before they took part.

## 2.2 Apparatus

### 2.2.1 All session types.

The sessions were implemented using SR Research Experiment Builder. The stimuli were presented to participants using Hyundai W 243s monitors with a 60 Hz refresh rate and a resolution of 1920 x 1200 pixels. Participants were seated 86 cm from the displays in a dimly-lit room and wore polarized spectacles in the multi-plane/single-plane condition, but wore no spectacles in the final transfer session when viewing the displays in 2D.

### 2.2.2 Non eye-tracking sessions.

Participants responded using the computer keyboard to indicate target-present and target-absent responses.

### 2.2.3 Eye-tracking sessions.

We recorded eye-movement behavior using an Eyelink 1000, set to record at a sampling rate of 1000 Hz (1 sample every millisecond). Recordings were taken from the right eye only. We used a nine-point calibration, which was accepted when the average error was below 0.5° of visual angle, and no single point exceeded an error of 1° of visual angle. Before each trial, a drift correct procedure was used, and calibrations were repeated when the error was above 1° of visual angle. Participants responded target-present/target-absent using a gamepad response box connected to the eye-tracking computer.

## 2.3 Stimuli

### 2.3.1 Display layout.

We generated the stimuli before data collection using customized code written in C# (Godwin, Holliman, Menneer, Liversedge, et al., 2015). The computer displays subtended 42.8° x 26.7° of visual angle. In order to prevent any problems caused by depth artefacts towards the edge of such large displays, images were presented within a central 28.5° x 21.4° region within the display, with the outer regions being left blank (white).

Within each trial, the region designated to contain objects was set out as a 4 x 3 grid of master grid cells. Each cell subtended 7.1° x 7.1° degrees of visual angle. The randomization procedure began by determining which of the grid cells objects should occupy. None of the objects within a grid cell could be placed on a cell boundary, ensuring objects in neighboring cells could not overlap. In order to make the displays appear more 'random' and less systematic, any occupied master grid cells were jittered by a random degree into neighboring master grid cells by a distance of up to 3.6° of visual angle (but only if the neighboring cells were unoccupied).

When a master grid cell was selected to contain objects, four objects were placed within the cell, one in each quadrant of the cell. The objects were then moved towards to the center of the master grid cell until each of them overlapped one another by the specified amount, which could be up to 90% of their visual area.

**2.3.2 Stimulus set.**

We used the same library of X-ray images as in our previous study (Godwin et al., 2017). Targets were designated as *metals* (guns/knives) or *Improvised Explosive Devices* (IEDs). Metals were predominantly blue in color; IEDs were primarily orange (the explosive component) along with blue shapes (metal wires, detonator, batteries, etc.). Distractors consisted of safe items typically found in baggage screening, including a variety of travel items such as sunglasses, keys, wallets, MP3 players, and so on. Object colors included blue, orange and black, depending upon the atomic density of the objects that had been imaged. Objects were presented at five different orientations, consisting of a canonical view, as well as rotations of 45º and 90º in the x- and y-planes. In total, a library of 270 target images and 962 distractor images was used.

**2.3.3 Implementation of depth.**

Objects were presented upon one of four stereoscopic depth planes (two appearing in front of the computer monitor; two appearing behind the computer monitor). The distances between neighboring planes were all equal. The perceived depth range for the monitor was around 12.5cm. In order to implement stereoscopic depth for the images, we began by generating a ‘main’ image for each trial, after which alternate rows of pixels were transposed to the left or right. When viewed while wearing polarized glasses, the alternate rows are presented to separate eyes, creating the percept of depth.

Participants viewed the displays under three different depth conditions (see Table 1). In the single-plane condition, all of the objects were presented on one of the four depth planes. The depth plane chosen for each trial was randomly selected under the constraint that an equal number of trials contained images presented at each depth plane. In the multi-plane condition, each object within a master grid cell was presented upon a different depth plane. As a consequence, this provided participants with depth cues to aid in the segmentation of objects in the multi-plane condition. Moreover, by using single-plane and multi-plane displays, we ensured that participants were always examining objects in depth, with the only difference being that the depth information was useful in one condition (multi-plane) but not in the other (single-plane). In the final session, objects were presented in 2D within the plane of the computer monitor.

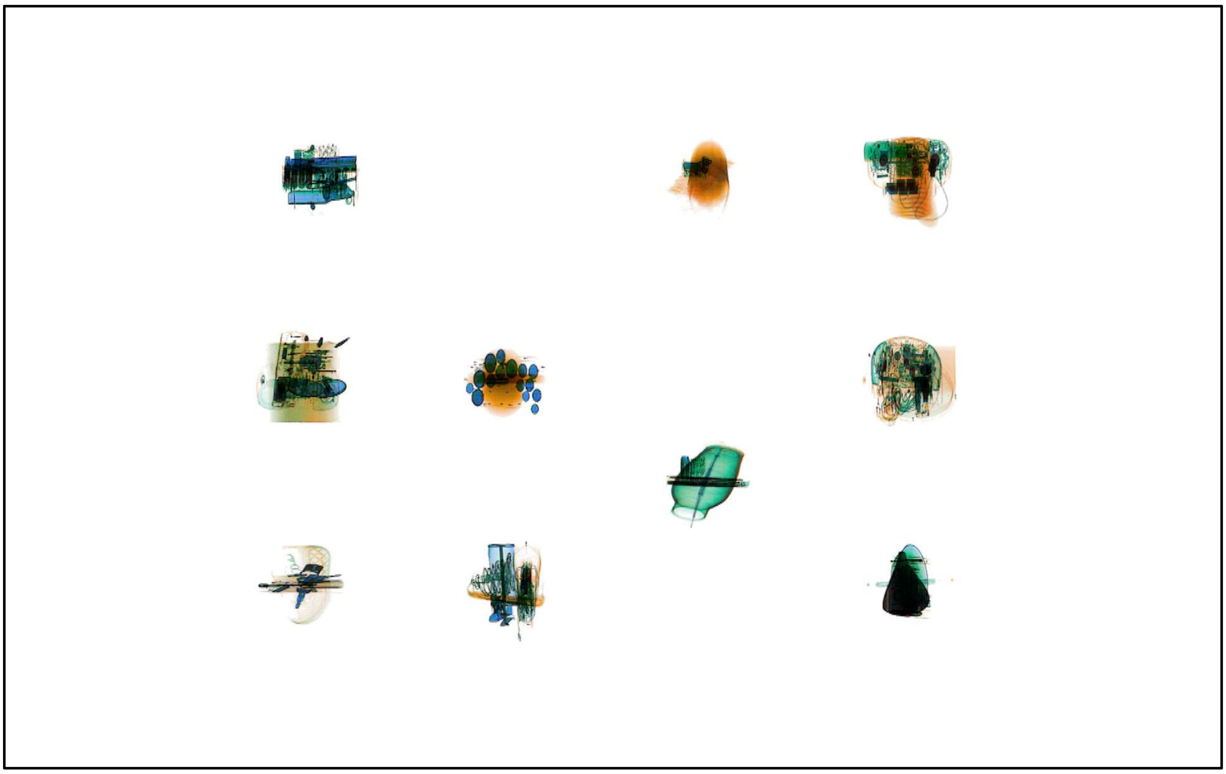

*Figure 1.* Example stimulus from the study (this version is in 2D).

**2.4 Design and Procedure**

Each participant engaged in 17 sessions of 288 trials (see Table 1), performing 4,896 trials in total. There were three blocks of trials within each session: single-target search for metals, single-target search for IEDs, and dual-target search. The order of the blocks was counterbalanced across participants, though each participant had the same order across all sessions. On each trial, participants were asked to respond as quickly and as accurately as possible whether they believed a target was present or absent. Following an incorrect response, a tone sounded from the computer. Before beginning the main study, participants completed a single session of searching the displays without overlap to enable them to gain experience with the task and the X-ray images.

Half of the participants were trained using multi-plane displays for the first 16 sessions; the remaining half of the participants were trained using single-plane displays for the first 16 sessions. Participants could take part in up to two testing sessions per day, with at least a one-hour break between sessions. Each testing session could last up to 90 minutes, with the entire study lasting up to 24 hours for each participant.

A single target was presented on 48 trials per block (50 % of trials). In half of the trials in each block, there was a setsize of 24 objects, and in the remaining half of trials in each block, there was a setsize of 40 objects. were preceded by a drift correct procedure (in eye-tracking sessions) or a fixation point (in non-eye-tracking sessions).

## 3.0 Results

### 3.1 Analytic Approach

We processed the raw data using the *eyeTrackR* package (Godwin & Muhl-Richardson, 2019) for R (R Development Core Team, 2013) . ANOVAs were conducted using the *ez* R package (Lawrence, 2015). Significant interactions were explored using further ANOVAs and Bonferroni-corrected *t*-tests where appropriate. We used generalized eta-squared (ges) to measure effect size (Bakeman, 2005). Across all of our analyses, measures that involved time were log-transformed prior to analysis, and proportion measures were arcsine-square-root transformed prior to analyses. However, untransformed means are presented in the figures.

We analyzed six measures in total, comprising two behavioral and four eye-movement measures. For the behavioral measures, we examined the standard measures of response accuracy and RTs. For the eye-movement measures, we separated the effects of perceptual selection and perceptual identification (Cain et al., 2013; Nodine & Kundel, 1987; Schwark et al., 2013). To examine perceptual selection, we analyzed the probability of fixating targets, and the time taken to fixate targets. To examine perceptual identification, we analyzed verification times, as well as the probability of identifying targets after having fixated them. The measures were averaged across the different set sizes to focus on the important results regarding depth and training effects.

For each measure, we conducted two initial mixed-design ANOVAs, one focusing on *training effects* and one focused on *transfer effects*. These enabled us to assess how training to search in the presence of depth influenced behavior and performance, and how that training transferred across to the final session wherein the displays were presented in 2D. The training effects ANOVAs included sessions 1-16 (or 4, 8, 12, and 16 for the eye-tracking analyses); the transfer effects ANOVAs included sessions 16 and 17 only (both of which involved eye-tracking). As such, it should be noted that the Session factor in the training and transfer ANOVAs refers to a differing number of sessions in each case.

The ANOVAs for the behavioral and eye-tracking measures shared the same basic design. Both the behavioral and the eye-tracking measures included factors of Depth (single-plane, multi-plane), Session (1-17) and Search Type (single-target, dual-target). The behavioral measures also included the Presence (target-present, target-absent) factor, which was not included in the eye-tracking measures since they were from target-present trials only.

Trials that had an RT of less than 200ms were removed as outliers, leading to the removal of approximately 1% of trials. Given the length of the study, a subset of the participants did not complete data collection, with 7 of the 44 participants being unable to take part in all sessions. The final dataset included data from 18 participants trained in multi-plane search and 19 participants trained in single-plane search. Finally, as a consequence of software and hardware errors in the eye-tracking system, there was a small degree of data loss. This resulting in the loss of 0.13% of eye-tracking trial data across all participants. Despite the data loss, no participants had empty cells during data analysis.

## 3.2 Behavioral Measures

**3.2.1 Response accuracy.** Descriptive statistics for the response accuracy data are presented in Figure 2, with ANOVA results for the response accuracy analyses presented in Table 2.

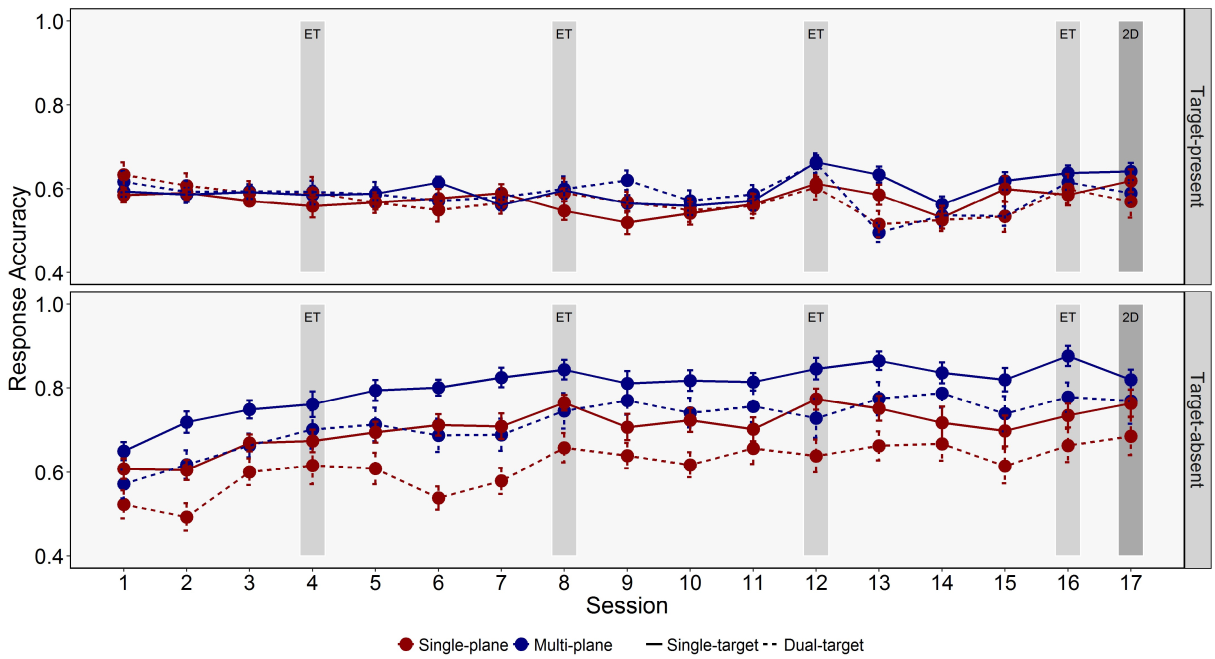


*Figure 2.* Response accuracy for the different sessions, levels of depth, search types, and target-present and target-absent trials. Error bars represent SE.

Table 2.

*Main Effects and Interactions from the ANOVAs examining Response Accuracy.*

| Model | Effect/Interaction | *F* | *df* | ges |
|---|---|---|---|---|
| Training Effects | Depth | 21.24*** | (1,35) | 0.06 |
| | Presence | 24.43*** | (1,35) | 0.22 |
| | Search Type | 330.59*** | (1,35) | 0.04 |
| | Session | 27.15*** | (15,525) | 0.04 |
| | Depth x Presence | 2.64 | (1,35) | 0.03 |
| | Depth x Search Type | 0.35 | (1,35) | 0.0000 |
| | Depth x Session | 2.56** | (15,525) | 0.004 |
| | Presence x Search Type | 27.69*** | (1,35) | 0.03 |
| | Presence x Session | 8.48*** | (15,525) | 0.06 |
| | Search Type x Session | 11.15*** | (15,525) | 0.01 |
| | Depth x Presence x Search Type | 0.35 | (1,35) | 0.0004 |
| | Depth x Presence x Session | 0.41 | (15,525) | 0.003 |
| | Depth x Search Type x Session | 1.18 | (15,525) | 0.001 |
| | Presence x Search Type x Session | 2.83*** | (15,525) | 0.01 |
| | Depth x Presence x Search Type x Session | 0.37 | (15,525) | 0.001 |
| Transfer Effects | Depth | 14.38*** | (1,35) | 0.05 |
| | Presence | 17.56*** | (1,35) | 0.24 |
| | Search Type | 56.7*** | (1,35) | 0.03 |
| | Session | 0.29 | (1,35) | 0.0002 |
| | Depth x Presence | 0.96 | (1,35) | 0.02 |
| | Depth x Search Type | 0.5 | (1,35) | 0.0003 |
| | Depth x Session | 6.4* | (1,35) | 0.004 |
| | Presence x Search Type | 2.73 | (1,35) | 0.01 |
| | Presence x Session | 0.01 | (1,35) | 0.0000 |
| | Search Type x Session | 1.35 | (1,35) | 0.001 |
| | Depth x Presence x Search Type | 0.11 | (1,35) | 0.0003 |
| | Depth x Presence x Session | 1.47 | (1,35) | 0.002 |
| | Depth x Search Type x Session | 3.26 | (1,35) | 0.002 |
| | Presence x Search Type x Session | 7.18* | (1,35) | 0.004 |
| | Depth x Presence x Search Type x Session | 0.19 | (1,35) | 0.0001 |

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

***Training effects.*** For the training effects ANOVA, response accuracy was higher in target-absent trials than target-present trials (Presence main effect), higher in single-target than dual-target search (Search Type main effect), and was higher in multi-plane than single-plane search (Depth main effect). There was also a main effect of session, indicating that, as would be expected, response accuracy improved as the sessions progressed. There were two key sets of interactions, namely between Depth and Session, and between Presence, Search Type and Session. We examine these in detail below.

The Depth x Session interaction was examined using independent *t*-tests focusing on Sessions 1 and 16. These revealed that although there was no Depth effect in the first session $t(35) = 1.6$, $p= .22$), a Depth effect had emerged by the final session ($t(35) = 4.68$, $p <. 0001$), demonstrating that training participants to search in multi-plane displays improved their response accuracy over time.

Next, to examine the Presence x Search Type x Session interaction, we examined target-present and target-absent trials separately using further ANOVAs. For both, there was a main effect of Session (*F*s > 4.2, *p*s < .001), reflecting an improvement in response accuracy over time, as well as interactions between Search Type and Session (*F*s > 3.29, *p*s < .0001). Subsequent *t*-tests failed to reveal Search Type effects for target-present trials in the first or final sessions (*t*s < 2.1, *p*s > .08), though response accuracy was lower in dual-target search than single-target search for absent trials (*t*s > 3.97, *p*s < .001).

***Transfer effects.*** The results that emerged for the transfer effects ANOVA mirrored those that emerged for the training effects ANOVA, with interactions between Depth and Session, as well as between Presence, Search Type and Session. We explored these in the same manner as for the training effects ANOVA, though did not repeat any *t*-tests for effects relating to session 16 within itself since these had already been conducted.

For the transfer session (session 17), as with the training sessions, response accuracy was higher for participants who had been trained in multi-plane than single-plane search ($t(35) = 2.41$, $p = .022$). This finding demonstrates that the benefits that emerged for the presence of depth during training transferred over to the search of two-dimensional displays.

Next, we again examined the Presence x Search Type x Session interaction by conducting additional ANOVAs for target-present and target-absent trials in sessions 16 and 17. These revealed a main effect of Search Type for target-absent trials ($F(1,35) = 16.67$, $p < .001$, ges = 0.06), demonstrating a dual-target cost for response accuracy in sessions 16 and 17. They also revealed an interaction between Search Type and Session for target-present trials ($F(1,35) = 10.38$, $p = .003$, ges = 0.01), with a subsequent *t*-test revealing that there was a dual-target cost for target-present trials in the final session ($t(36) = 3.3$, $p = .002$), with response accuracy being lower in dual-target than single-target search.

***Summary.*** The response accuracy analyses replicate the results of our previous study (Godwin et al., 2017), wherein searching in multi-plane displays improved accuracy for overlapping X-ray stimuli. Importantly, this effect also transferred across to confer a benefit to response accuracy even when searching 2D displays after having been trained during search of multi-plane displays. These benefits emerged for both target-present and target-absent trials.

**3.2.2 Response times.** Descriptive statistics for the response time data are presented in Figure 3, with ANOVA results for the response times presented in Table 3.

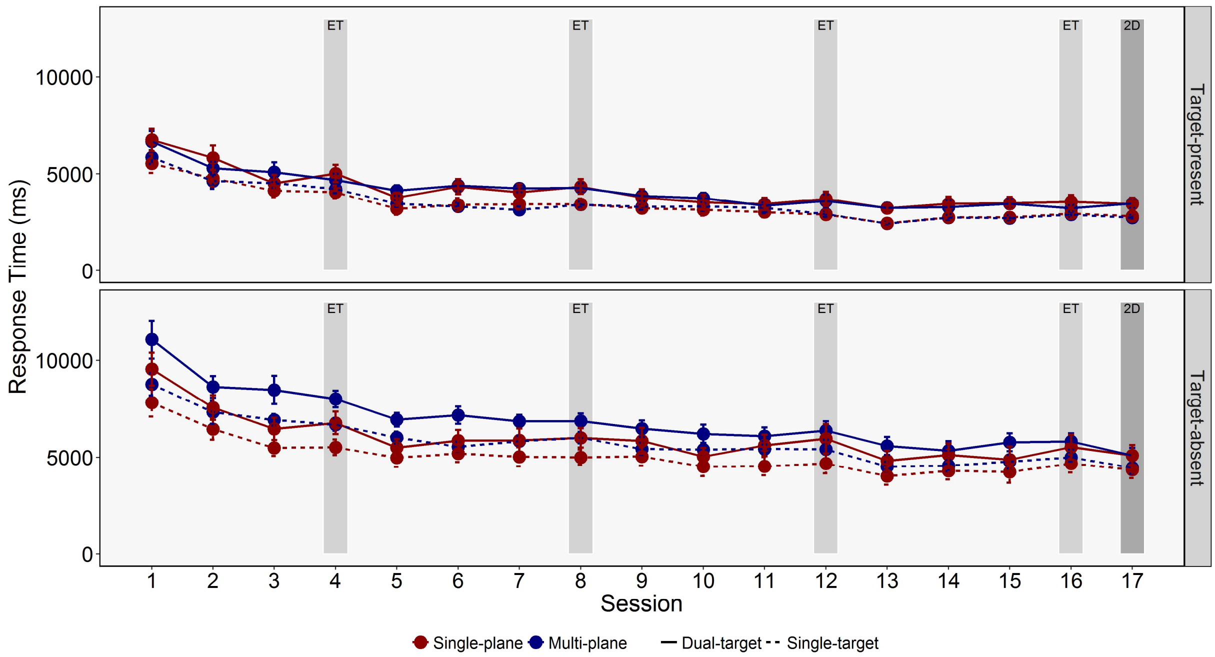


*Figure 3.* Response Times for the different Sessions, levels of Depth, Search Types, and Target-present and Target-absent Trials. Error bars represent SE.

Table 3.

*Main Effects and Interactions from the ANOVAs examining Response Times*

| Model | Effect/Interaction | *F* | *df* | ges |
|---|---|---|---|---|
| Training Effects | Depth | 1.76 | (1,35) | 0.03 |
| | Presence | 494.41*** | (1,35) | 0.42 |
| | Search Type | 229.44*** | (1,35) | 0.07 |
| | Session | 64.92*** | (15,525) | 0.24 |
| | Depth x Presence | 14.27*** | (1,35) | 0.02 |
| | Depth x Search Type | 0.04 | (1,35) | 0.0000 |
| | Depth x Session | 0.8 | (15,525) | 0.004 |
| | Presence x Search Type | 1.86 | (1,35) | 0.0001 |
| | Presence x Session | 5.73*** | (15,525) | 0.004 |
| | Search Type x Session | 2.74*** | (15,525) | 0.004 |

| | | | | |
|---|---|---|---|---|
| | Depth x Presence x Search Type | 0.81 | (1,35) | 0.0000 |
| | Depth x Presence x Session | 0.93 | (15,525) | 0.001 |
| | Depth x Search Type x Session | 0.75 | (15,525) | 0.001 |
| | Presence x Search Type x Session | 7.18*** | (15,525) | 0.003 |
| | Depth x Presence x Search Type x Session | 1.01 | (15,525) | 0.0005 |
| Transfer Effects | Depth | 0.32 | (1,35) | 0.01 |
| | Presence | 319.26*** | (1,35) | 0.43 |
| | Search Type | 84.95*** | (1,35) | 0.06 |
| | Session | 4.4* | (1,35) | 0.01 |
| | Depth x Presence | 2.67 | (1,35) | 0.01 |
| | Depth x Search Type | 0 | (1,35) | 0.0000 |
| | Depth x Session | 0.01 | (1,35) | 0.0000 |
| | Presence x Search Type | 3.92 | (1,35) | 0.001 |
| | Presence x Session | 26.91*** | (1,35) | 0.01 |
| | Search Type x Session | 1.5 | (1,35) | 0.0005 |
| | Depth x Presence x Search Type | 0.03 | (1,35) | 0.0000 |
| | Depth x Presence x Session | 6.64* | (1,35) | 0.002 |
| | Depth x Search Type x Session | 0.91 | (1,35) | 0.0003 |
| | Presence x Search Type x Session | 2.97 | (1,35) | 0.001 |
| | Depth x Presence x Search Type x Session | 0.99 | (1,35) | 0.0003 |

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

***Training effects.*** For the ANOVA conducted upon sessions 1-16, participants were slower to response in target-absent than target-present trials (Presence main effect), were slower to respond in dual-target than single-target search (Search Type main effect), and responded more rapidly as the sessions progressed (Session main effect). There were two key sets of interactions that we explored in detail: Depth x Presence and Presence x Search Type x Session.

For the Depth x Presence interaction, we conducted further ANOVAs for target-present and target-absent trials separately. These revealed that RTs in target-absent trials were longer for multi-plane than single-plane search ($F$(1,35) = 4.8, $p$ = .041, ges = .085), though this was not the case for target-present trials ($F < 1$). This is an important result and one that we shall return to in more detail in the Discussion.

Turning to the Presence x Search Type x Session interaction, we again broke the analyses down with further ANOVAs conducted upon target-present and target-absent trials separately. For target-absent trials, there RTs were longer in dual-target than single-target search, and there was a reduction in RTs for target-absent trials between sessions 1 and 16 ($Fs > 43$, $p$s < .0001). For target-present trials, there was an interaction between Search Type and Session ($F(15,525) = 7.07$, $p < .0001$, ges = .015). Subsequent $t$-tests revealed that target-present RTs were longer in dual-target search than single-target search in both sessions 1 and 16 ($t$s > 2.8, $p$s < .05). Overall, therefore, RTs were subject to a dual-target cost that was not eliminated over time.

***Transfer effects.*** The transfer effects ANOVA for RTs revealed one key interaction only, namely between Depth, Presence and Session. Conducting further ANOVAs for target-present and target-absent trials separately failed to reveal effects of Depth ($Fs < 1$), though RTs did reduce between sessions 16 and 17 for target-absent trials only ($F(1,35) = 13.83$, $p = .0006$, ges = .019).

***Summary.*** In our previous study, we found no evidence that RTs were directly influenced by the presence of depth in the displays (Godwin et al., 2017), yet here, RTs for target-absent trials were longer for multi-plane than single-pane searches. Although this difference was not reflected in the transfer session, the main finding shall be returned to in the Discussion wherein we will draw out comparisons and explanations for these differing results to our previous work.

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

We selected key eye movement measures to focus on how and why participants failed to detect targets when searching (Cain et al., 2013; Nodine & Kundel, 1987; Schwark et al., 2013). These measures involved failures of perceptual selection (i.e., time to fixate targets, and the probability of fixating targets), alongside failures of perceptual identification (verification time and the probability of identifying targets after fixating them). Given that the behavioral analyses demonstrated that training in multi-plane displays primarily influenced response accuracy rates, and only influenced RTs for target-present trials, we anticipated that the measures of time in this regard (time to fixate targets, verification time) would be unlikely to show effects of depth.

**3.3.1 Time to fixate targets.** Descriptive statistics for the times to fixate targets are presented in Figure 4, with ANOVA results for the times to fixate targets analyses presented in Table 4.

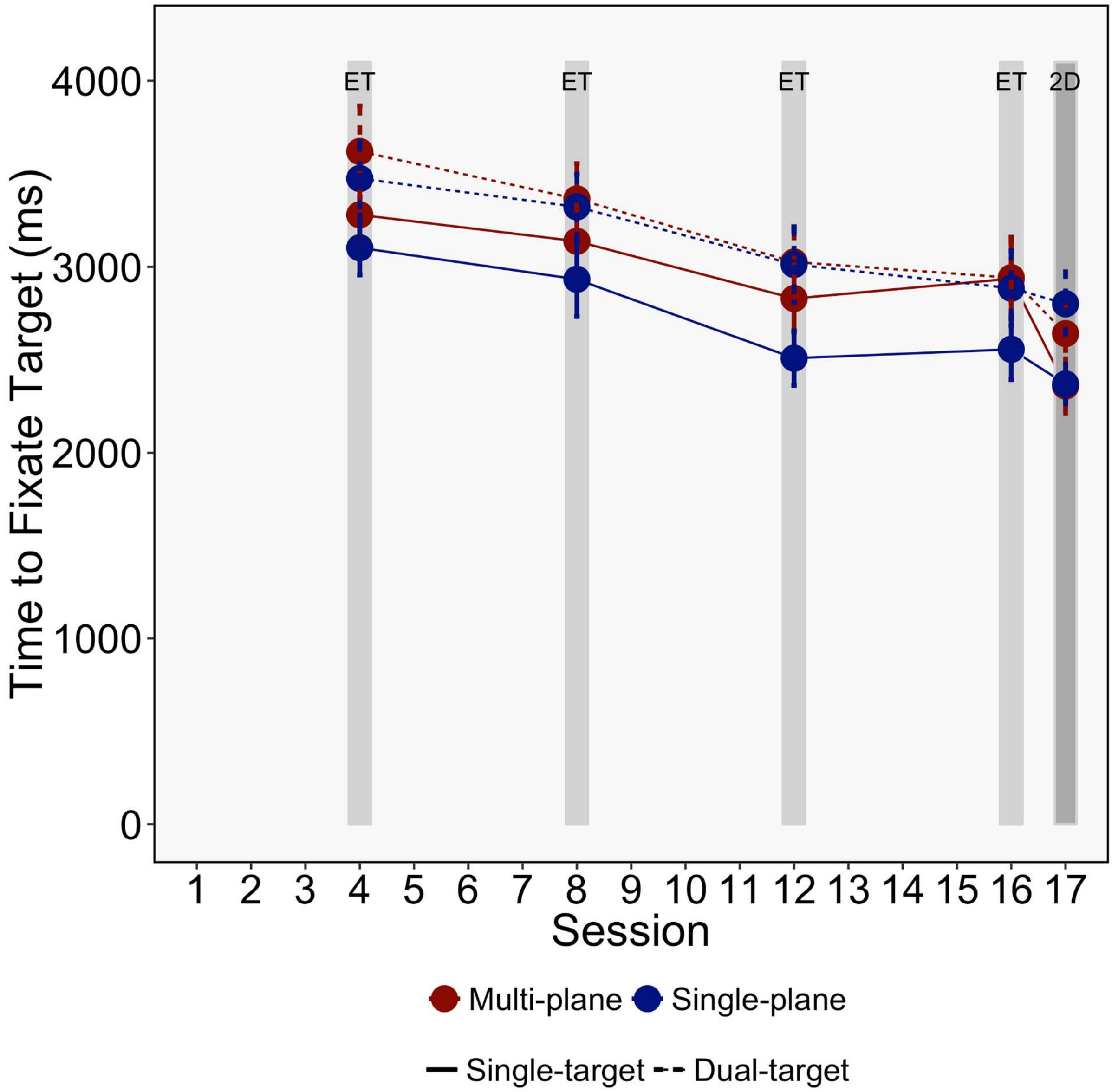


*Figure 4.* Time to fixate the targets for the different Sessions, levels of depth and search types. Error bars represent SE.

Table 4.

*Main Effects and Interactions from the ANOVAs examining the Time to Fixate Targets.*

| Model | Effect/Interaction | *F* | *df* | ges |
|---|---|---|---|---|
| Training Effects | Depth | 1.04 | (1,35) | 0.02 |
| | Search Type | 39.45*** | (1,35) | 0.07 |
| | Session | 57.32*** | (3,105) | 0.21 |
| | Depth x Search Type | 0.18 | (1,35) | 0.0003 |
| | Depth x Session | 0.81 | (3,105) | 0.004 |
| | Search Type x Session | 2.89* | (3,105) | 0.01 |
| | Depth x Search Type x Session | 0.27 | (3,105) | 0.001 |
| Transfer Effects | Depth | 0.07 | (1,35) | 0.002 |
| | Search Type | 120.53*** | (1,35) | 0.14 |
| | Session | 0.04 | (1,35) | 0.0001 |
| | Depth x Search Type | 0.52 | (1,35) | 0.001 |
| | Depth x Session | 0.36 | (1,35) | 0.001 |
| | Search Type x Session | 3.53 | (1,35) | 0.01 |
| | Depth x Search Type x Session | 0 | (1,35) | 0.0000 |

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

***Training effects.*** The training effects ANOVA that examined the time taken to fixate targets showed that participants were slower to fixate targets in dual-target than single-target search (main effect of Search Type), consistent with previous research showing that dual-target search impairs guidance processes (Stroud et al., 2012). Participants also reduced their time taken to fixate targets as the sessions progressed (main effect of Session). There was an interaction between Session and Search Type, and subsequent *t*-tests revealed that the dual-target cost for time to fixate targets was present in both sessions 4 and 16 ($t$s > 3, $p$s < .01).

***Transfer effects.*** The transfer effects ANOVA for the time to fixate targets revealed a main effect of Search Type only, demonstrating that the time to fixate targets was longer for dual-target than single-target searches.

***Summary.*** As with our previous study (Godwin et al., 2017), the presence of depth in the displays did not influence the time taken to fixate targets. However, dual-target search increased the time taken to fixate targets. This was the case both during the training and transfer sessions. As noted above, it is perhaps not surprising that depth did not influence the time to fixate targets since the RTs for target-present trials were also not influenced by the presence of depth in the displays.

**3.3.2 Probability of fixating targets.** Descriptive statistics for the probability of fixating targets are presented in Figure 5, with ANOVA results for the probability of fixating targets analyses presented in Table 5.

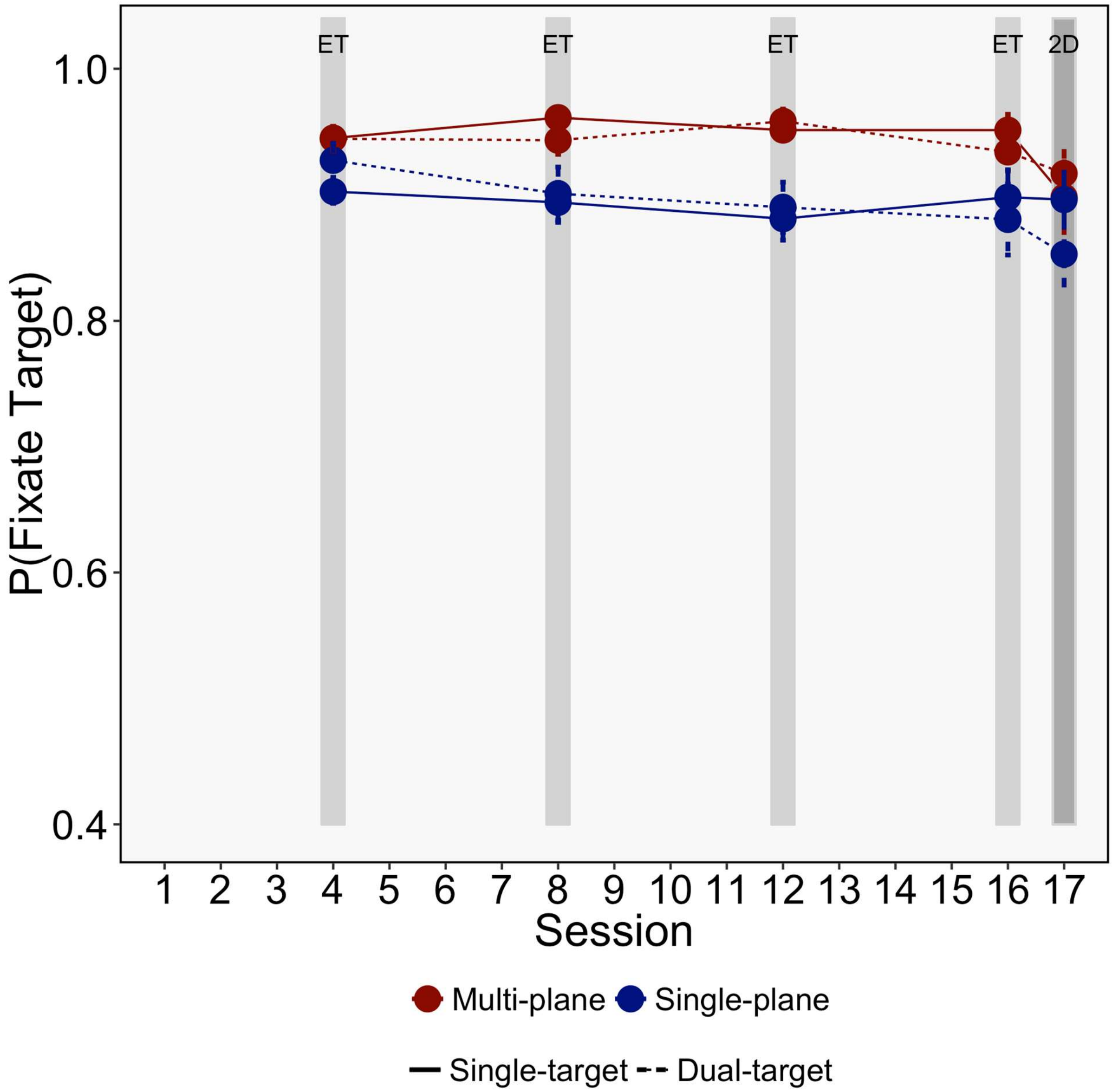


*Figure 5.* Probability of fixating the target for the different Sessions, levels of depth and search types. Error bars represent SE.

Table 5.

*Main Effects and Interactions from the ANOVAs examining the Probability of Fixating Targets.*

| Model | Effect/Interaction | *F* | *df* | ges |
|---|---|---|---|---|
| Training Effects | Depth | 7.73** | (1,35) | 0.13 |
| | Search Type | 0.01 | (1,35) | 0.0000 |
| | Session | 1.43 | (3,105) | 0.01 |
| | Depth x Search Type | 1.68 | (1,35) | 0.002 |
| | Depth x Session | 2.54 | (3,105) | 0.01 |
| | Search Type x Session | 3.27* | (3,105) | 0.01 |
| | Depth x Search Type x Session | 0.88 | (3,105) | 0.002 |
| Transfer Effects | Depth | 2.33 | (1,35) | 0.05 |
| | Search Type | 2.52 | (1,35) | 0.01 |
| | Session | 6.69* | (1,35) | 0.02 |
| | Depth x Search Type | 2.63 | (1,35) | 0.01 |
| | Depth x Session | 1.15 | (1,35) | 0.003 |
| | Search Type x Session | 0.06 | (1,35) | 0.0001 |
| | Depth x Search Type x Session | 2.77 | (1,35) | 0.01 |

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

***Training effects.*** The initial ANOVA examining training effects for the probability of fixating targets revealed a number of main effects and interactions. There was a main effect of Depth, indicating that participants were more likely to fixate targets in multi-plane than single-plane search. There was also an interaction between Search Type and Session. In order to examine the Search Type x Session interaction, we conducted $t$-tests comparing the multi-plane and single-plane search performance in sessions 4 and 16. Surprisingly, these failed to reach significance ($t$s < 1.2, $p$s > .5).

***Transfer effects.*** The transfer effects ANOVA for the probability of fixating targets revealed an effect of Session only, with participants being less likely to fixate targets in the transfer session.

***Summary.*** In line with our previous study (Godwin et al., 2017), we again found that participants were more likely to fixate targets in multi-plane than single-plane search. This was true, however, only in the training sessions, but not in the transfer session. The increase in the probability of fixating targets could explain why response accuracy was higher in multi-plane than single-plane search: in complex search tasks of this type, fixating the target is an important prerequisite for target detection.

**3.3.3 Verification time.** Descriptive statistics for verification times are presented in Figure 6, with ANOVA results for verification times presented in Table 6.

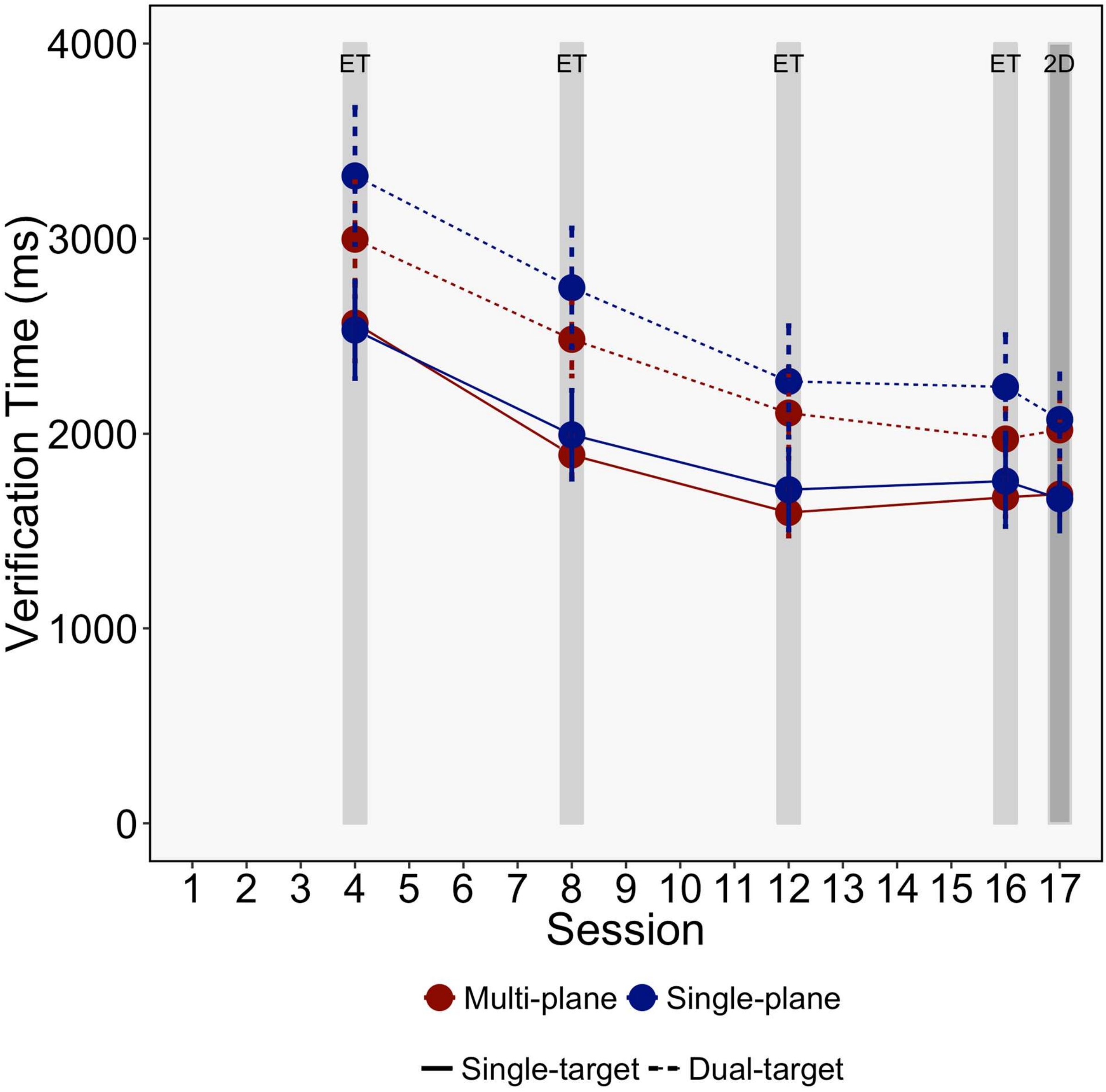


*Figure 6.* Verification time for the different sessions, levels of depth and search types. Error bars represent SE.

Table 6.

*Main Effects and Interactions from the ANOVAs examining Verification Time.*

| Model | Effect/Interaction | *F* | *df* | ges |
|---|---|---|---|---|
| Training Effects | Depth | 0.17 | (1,35) | 0.004 |
| | Search Type | 80.8*** | (1,35) | 0.10 |
| | Session | 75.66*** | (3,105) | 0.21 |
| | Depth x Search Type | 0.7 | (1,35) | 0.001 |
| | Depth x Session | 0.47 | (3,105) | 0.002 |
| | Search Type x Session | 2.85* | (3,105) | 0.004 |
| | Depth x Search Type x Session | 0.74 | (3,105) | 0.001 |
| Transfer Effects | Depth | 0.01 | (1,35) | 0.0002 |
| | Search Type | 48.53*** | (1,35) | 0.07 |
| | Session | 0.06 | (1,35) | 0.0001 |
| | Depth x Search Type | 1.15 | (1,35) | 0.002 |
| | Depth x Session | 2.22 | (1,35) | 0.003 |
| | Search Type x Session | 1.47 | (1,35) | 0.001 |
| | Depth x Search Type x Session | 0.25 | (1,35) | 0.0002 |

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

***Training effects.*** The training effects ANOVA for verification time revealed that participants were slower to verify targets in dual-target search than single-target search (main effect of Search Type), and that verification times reduced as the sessions progressed (main effect of Session). There was also a Search Type x Session interaction, which was examined using *t*-tests. These indicated that verification times were slower for dual-target than single-target search in both sessions 4 and 16 (*t*s > 4.3, *p*s < .0001).

***Transfer effects.*** The transfer effects ANOVA for verification times revealed an effect of Search Type only, indicating that participants were slower to verify targets in dual-target search than single-target search, mirroring the findings of the training effects analyses.

***Summary.*** Again the results for verification times were in line with our previous study (Godwin et al., 2017): the presence of depth in the displays did not influence verification times, suggesting that presenting objects on different depth planes does not aid in object identification processes relative to presenting them on the same depth plane. This was as expected given the lack of depth effects for the target-present trial RTs. However, participants clearly were learning to better identify the targets as the sessions progressed, as verification times reduced across the sessions. This could, at least in part, account for the reduction in RTs for later sessions.

**3.3.4 Probability of identifying targets after fixating them.** Descriptive statistics for the probability of identifying targets after fixating them are presented in Figure 7, with ANOVA results presented in Table 7.

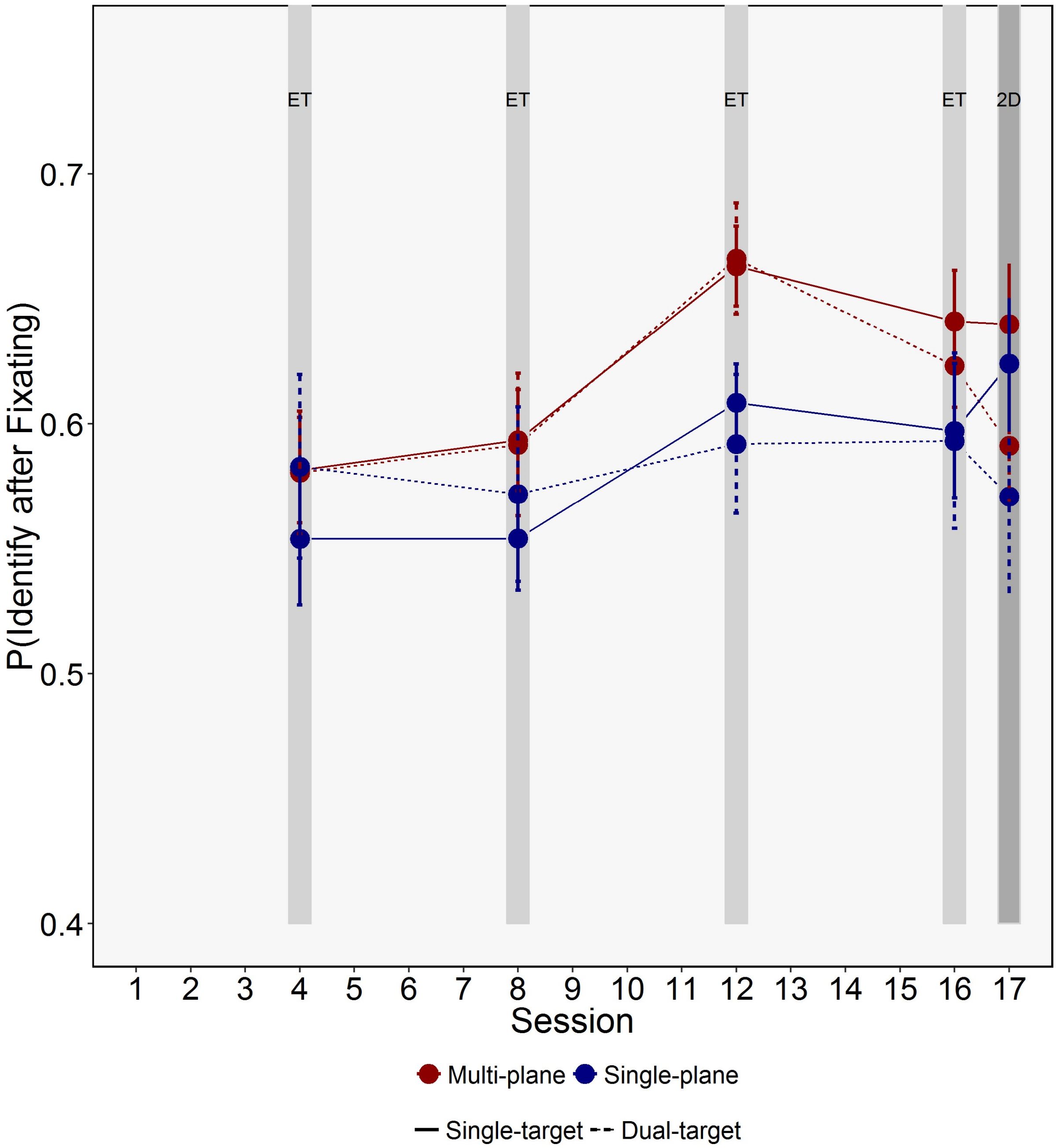


*Figure 7.* Probability of identifying targets after fixating them for the different sessions, levels of depth and search types. Error bars represent SE.

Table 7.

*Main Effects and Interactions from the ANOVAs examining the Probability of Identifying Targets after Fixating them.*

| Model | Effect/Interaction | *F* | *df* | ges |
|---|---|---|---|---|
| Training Effects | Depth | 1.9 | (1,35) | 0.03 |
| | Search Type | 0.02 | (1,35) | 0.0000 |
| | Session | 6.06*** | (3,105) | 0.05 |
| | Depth x Search Type | 0.32 | (1,35) | 0.001 |
| | Depth x Session | 0.91 | (3,105) | 0.01 |
| | Search Type x Session | 0.67 | (3,105) | 0.002 |
| | Depth x Search Type x Session | 0.52 | (3,105) | 0.002 |
| Transfer Effects | Depth | 0.66 | (1,35) | 0.01 |
| | Search Type | 5.04* | (1,35) | 0.02 |
| | Session | 0.37 | (1,35) | 0.001 |
| | Depth x Search Type | 0.03 | (1,35) | 0.0001 |
| | Depth x Session | 0.71 | (1,35) | 0.002 |
| | Search Type x Session | 5.46* | (1,35) | 0.01 |
| | Depth x Search Type x Session | 0.29 | (1,35) | 0.0004 |

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

***Training effects.*** The ANOVA examining training effects for the probability of identifying targets after fixating them revealed an effect of Session only, with participants being more likely to fixate and then identify the targets as the sessions progressed.

***Transfer effects.*** The ANOVA that examined transfer effects for the probability of identifying targets after fixating them revealed an effect of Search Type, as well as an interaction between Search Type and Session. A subsequent *t*-test revealed that there were no differences between single-target and dual-target search in session 16 ($t < 1$), though participants were less likely to identify targets after fixating them in dual-target than single-target search during the transfer session ($t(36)= 2.96$, $p = .01$).

***Summary.*** In a pattern of results that was not consistent with our previous study (Godwin et al., 2017), here we found that the presence of depth in the displays did not influence the probability of identifying fixated targets. Previously, we found that the presence of depth in the displays increased the probability of identifying targets after fixating them. Within the context of the present study, however, these findings align with those for verification times, suggesting that object identification processes were uninfluenced by the presence of depth. However, participants did clearly learn to better identify targets as the sessions progressed.

## 4.0 Discussion

Standard visual search experiments involve asking participants to search for a single target in displays wherein objects do not overlap with one another, and are all presented on a single depth plane (Eckstein, 2011). Though such an approach has of course been invaluable in studying search, there is a surprising lack of data and models relating to searching complex, overlapping displays, as well as real-world tasks such as physically searching the environment for targets (Riggs et al., 2017, 2018; Smith, Wallace, Hood, & Gilchrist, 2009). The present study builds upon recent work targeted at better understanding search through overlapping displays in depth, focusing upon training and transfer effects that arise from searching overlapping displays containing depth. To date, there is mixed evidence in terms of whether depth aids performance in training (McIntire, Havig, & Geiselman, 2014), and, as a result, there appears a need for further investigation of training in depth. This is true for both single- and dual-target searchers, where it is important to determine whether the presence of depth in the displays can help ameliorate known costs to search performance that arise in dual-target search.

We previously studied search of overlapping displays (Godwin et al., 2017), and asked the question: does presenting overlapping objects upon different depth planes aid in resolving the problems associated with overlap? We found that depth was beneficial to search performance through overlapping displays via an increase in response accuracy (but not in RTs), but only for some stimulus types (transparent polygons, X-ray objects, and target-present trials for real-world objects). Our pattern of results, we argued, was consistent with the idea that depth facilitated object segmentation processes in search through overlapping displays by providing an additional cue to object identity. The eye movement analyses found evidence that the presence of depth attenuated the effects of overlap upon the probability of fixating targets, and also increased the probability that participants would identify targets after fixating them. This was the case across all stimulus types. Our previous study thus provided evidence that the presence of depth in the displays aided perceptual selection and perceptual identification processes. Either or both of these could explain the increase in response accuracy that we observed previously.

We therefore had two objectives in the present study. First, we sought to replicate our previous pattern of results. Second, we sought to determine whether the benefits conferred by experience of searching displays containing overlapping objects presented on different depth planes could transfer across to searching in two-dimensional displays. If that is the case, then adding stereoscopic depth to search displays could be used as a training tool to encourage novice searchers to adopt new strategies for searching and interpreting displays when depth information must be inferred from two-dimensional pictorial cues alone. We therefore engaged participants in a series of seventeen sessions of single- and dual-target searches. Half of the participants were trained using single-plane displays; half were trained using multi-plane displays. The final session involved presenting two-dimensional displays to participants to test for transfer effects.

Focusing on the training sessions to begin with, our results replicated those that we obtained previously (Godwin et al., 2017). As in our previous study, the presence of depth improved response accuracy. In the eye movement analyses, again, as in our previous study, the presence of depth increased the probability that participants would fixate targets, but did not influence the time to fixate targets, nor did it influence the verification time. Generally speaking, our results therefore mirrored those found previously with these stimuli.

There were, however, some differences between the present set of results and our previous study. Here, the presence of depth increased RTs on target-absent trials, which was not the case previously. In addition, unlike our previous study, the presence of depth did not influence the probability of identifying targets after having fixated them. Fortunately, there is a simple explanation for this. The most likely reason for divergence in the pattern of results obtained here beyond our previous study are that the present study had substantially increased power, and this arose through a number of different routes. First, there were more sessions/trials here compared with our previous study. Second, the displays used here had only the highest level of overlap (90%), rather than different levels of overlap as used previously (where we used 0%, 45% and 90% overlap). Third, every object in the displays used here overlapped with other objects; even in the 90% overlap condition in our previous study, only half of the objects in each display overlapped with other objects. Therefore, put simply, the present study had a substantially higher level of power due to increased trials, higher levels of overlap, and by having more objects overlapping one another. If anything, this suggests that our previous study may have *underestimated* the effects that searching in depth have upon performance, and this is particularly important for real-world tasks wherein the majority of objects have a high degree of overlap with other objects.

In terms of transfer effects, the pattern of effects was somewhat weaker than those that emerged during training. To some extent, this reduction might have been expected given that there was only a single transfer session, and that participants had not searched two-dimensional X-ray displays until that point. Despite the reduced effects, we did find that, in the transfer session, participants who had been trained using multi-plane displays did exhibit higher response accuracy than those who had been trained using single-plane displays. This result alone indicates that training in depth does indeed confer benefits to search performance that transfer to two-dimensional displays. Of course, further work is required to determine the longevity of these transfer effects.

The final aspect of the study that we examined was comparing dual-target and single-target searches. Previous work has shown that, when searching for two targets, search is both less accurate and less rapid than when searching for a single target. Here, we sought to address whether the presence of depth in the displays might ameliorate these costs to performance when searching. Throughout the analyses, we found no clear evidence that the presence of depth reduced the differences (or lack thereof) between single-target and dual-target searches in any of our measures. Still, the presence of depth in the displays, as discussed above, did confer *overall* benefits to performance, but did not *specifically aid* in reducing the costs associated with searching for two targets simultaneously.

Viewing the results from a broader perspective, we have now conducted two sets of studies that examined the effects that depth have upon search performance. What we can say is that the presence of depth appears to influence response accuracy and also influence the probability of fixating targets. The question of whether depth influences RTs and/or the probability of identifying targets after fixating them is less clear-cut, since the results diverged between our two studies on these measures. However, if anything, the results of the present study have a higher degree of power, so it may simply be the case that the effects of depth are quite subtle and/or only emerge when overlap levels are very high indeed.

Drawing the findings together, it appears that searching in the presence of depth, and being trained to search in the presence of depth, encouraged participants to engage in a more exhaustive search. Exhaustiveness is an important part of visual search, particularly in complex displays in which all objects need to be carefully examined to determine whether they are the target (Chun & Wolfe, 1996). Indeed, when participants quit searching too soon, they are highly likely to miss targets (Rich, Hidalgo-sotelo, Kunar, Wert, & Wolfe, 2005). Later quitting also reduces the chance that participants will make a false alarm, since false alarms are often triggered by a 'guess' when searching has proceeded for a long period of time (Chun & Wolfe, 1996).

If participants here were more exhaustive in their searches of multi-plane than single plane displays, then this could explain why the probability of fixating targets was higher in multi-plane than single-plane displays but the time to fixate targets did not differ. Under this explanation, response accuracy (increased hit rate, reduced false alarm rate) was higher for participants searching in depth as they obtained more information regarding object identities than those examining displays that did not contain depth. The increased exhaustiveness did not reach significance for target-present trials, however, most likely because targets are found on present-trials, on average, after only half of the objects have been examined. Evidence of search exhaustiveness should, therefore, be more easily demonstrated in target-absent than target-present trials.

With that possibility in mind, we conducted a further brief analysis of the data from this study, examining the proportion of objects fixated in each trial using an ANOVA (of the same design as those used for the 'training' analyses with the exception that it focused on target-absent trials only). This ANOVA revealed a main effect of Depth ($F(1,35) = 6.95$, $p = 0.012$), and a Depth x Search Type interaction, $F(1,35)= 14.78$, $p < .001$. Subsequent *t*-tests revealed evidence of an effect of Depth for single-target search only ($t(35) = 3.29$, $p < .01$), with participants fixating more objects in multi-plane ($M = 0.89$, $SE = 0.01$) than single-plane searches ($M = 0.81$, $SE = 0.01$). For dual-target search, the comparison did not reach significance ($t(35) = 1.95$, $p = .12$), perhaps due to a ceiling effect in how many objects participants examined in these conditions. These analyses provide further evidence to suggest that in real-world searches containing displays with a high degree of overlapping transparent objects, sometimes the only route to improving response accuracy may be to engage in a more exhaustive (or more detailed) search. Indeed, it is interesting to note that the strategic shift in search behavior that arises from searching in the presence of depth shares similarities with how expert searchers examine displays of this type (e.g. Biggs & Mitroff, 2014).

Depth, as we have studied it here, encouraged participants to engage in an exhaustive strategy, and we have some evidence that this strategy transferred over to searching of two-dimensional displays. How is it that the presence of depth can train participants to engage in such a strategy? The results are consistent with the presence of depth allowing information to be available in the display that is not available (or difficult to access) in single-plane images. Search becomes longer in order to give enough time to access that information, and, once that information is accessed, it allows more accurate responses. The source of the extra information in depth seems likely to be from the segregation of overlapping objects in the display, given there is no other qualitative difference between multi-plane and single-plane images. It also seems that separation in depth allows a participant to learn how overlapping objects interact and combine in 2D images such that the increase in accuracy transfers from multi-plane to 2D images.

Our results are important for practitioners in real-world tasks, as it opens up the possibility of training searchers using depth information in complex, overlapping displays, to maximize their performance wherever possible, potentially not just using depth, but other training regimes that encourage participants to search in an exhaustive fashion. Moreover, compared alongside our previous study, we have begun to map out the points at which the presence of depth is beneficial to search, given that the benefits depend to a significant extent upon the stimuli being searched, and their degree of overlap.

Before closing, it is important to make one final point for consideration. During everyday visual searches, the presence of depth and overlap are the norm rather than the exception. As a consequence, we are all, to some extent, ‘experts’ at searching environments containing overlapping objects on different depth planes. Standard search studies, however, involve the presentation of objects on a single depth plane to participants. It is therefore possible that our results presented here, and in other studies that have examined depth, have reversed the true situation. Rather than framing the analyses in terms of how the addition of depth influences behavior and performance, it may be more precise to frame the analyses in terms of how the *removal* of depth (such as in single-plane search or standard search tasks) influences behavior and performance. Put another way, depth may not actually improve search performance, but rather the lack of depth may impair search performance.

*Acknowledgements:* This work was supported by funding from the Economic and Social Sciences Research Council (grant ref. ES/I032398/1). The authors would like to thank Stuart Pugh and Patrick Clark for their assistance with data collection.

# References

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*(3), 379–384. https://doi.org/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. https://doi.org/10.1371/journal.pone.0086848

Biggs, A. T., & Mitroff, S. R. (2014). Differences in multiple-target visual search performance between non-professional and professional searchers due to decision-making criteria. *British Journal of Psychology*, 1–13. https://doi.org/10.1111/bjop.12096

Buser, D., Sterchi, Y., & Schwaninger, A. (2020). Why stop after 20 minutes? Breaks and target prevalence in a 60-minute X-ray baggage screening task. *International Journal of Industrial Ergonomics*, *76*, 102897. https://doi.org/10.1016/j.ergon.2019.102897

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. https://doi.org/10.1080/13506285.2013.843627

Chun, M. M., & Wolfe, J. M. (1996). Just say no: how are visual searches terminated when there is no target present? *Cognitive Psychology*, *30*(1), 39–78. https://doi.org/10.1006/cogp.1996.0002

Donnelly, N., Muhl-richardson, A., Godwin, H. J., & Cave, K. R. (2019). Using Eye Movements to Understand how Security Screeners Search for Threats in X-ray Baggage. *Vision*, (X), 1–15.

Eckstein, M. P. (2011). Visual search : A retrospective. *Journal of Vision*, *11*(5:14), 1–36. https://doi.org/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. https://doi.org/10.1167/13.8.11

Godwin, H. J., Holliman, N. S., Menneer, T., Liversedge, S. P., Cave, K. R., & Donnelly, N. (2015). 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., Liversedge, S. P., Kirkby, J. A., Boardman, M., Cornes, K., & Donnelly, N. (2015). The influence of experience upon information-sampling and decision-making behaviour during risk assessment in military personnel. *Visual Cognition*, *23*(4), 415–431. https://doi.org/10.1080/13506285.2015.1030488

Godwin, H. J., Menneer, T., Liversedge, S. P., Cave, K. R., Holliman, N. S., & Donnelly, N. (2017). Adding Depth to Overlapping Displays can Improve Visual Search Performance. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(8), 1532–1549. https://doi.org/10.1037/xhp0000353

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), 150–159. https://doi.org/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*. https://doi.org/10.3758/s13423-015-0970-9

Godwin, H. J., & Muhl-Richardson, A. (2019). eyeTrackR: Organising and Analysing Eye-Tracking Data.

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*.

He, Z., & Nakayama, K. (1992). Surfaces versus features in visual search. *Nature*, *359*, 231–233. https://doi.org/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), 158–170. https://doi.org/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*. https://doi.org/10.1037/xhp0000053

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.

McIntire, J. P., Havig, P. R., & Geiselman, E. E. (2014). Stereoscopic 3D Displays and Human Performance: A Comprehensive Review. *Displays*, *35*(1), 18–26. https://doi.org/10.1016/j.displa.2013.10.004

McSorley, E., & Findlay, J. M. (2001). Visual search in depth. *Vision Research*, *41*(25–26), 3487–3496. https://doi.org/10.1016/S0042-6989(01)00197-3

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. https://doi.org/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. https://doi.org/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), 404–418. https://doi.org/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), 381–390. https://doi.org/10.3758/BF03206781

Muhl-Richardson, A., Godwin, H. J., Garner, M., Hadwin, J. A., Liversedge, S. P., & Donnelly, N. (2018). Individual Differences in Search and Monitoring for Color Targets in Dynamic Visual Displays. *Journal of Experimental Psychology: Applied*. https://doi.org/10.1037/xap0000155

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. https://doi.org/10.1068/p180055

Nakayama, K., & Silverman, G. H. (1986). Serial and Parallel Processing of Visual Feature Conjunctions. *Nature*, *320*(6059), 264–265. https://doi.org/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), 202–218. https://doi.org/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.

Rich, A. N., Hidalgo-sotelo, B., Kunar, M. A., Wert, M. J. Van, & Wolfe, J. M. (2005). What happens during search for rare targets ? Eye movements in low prevalence visual search, (2004), 2005.

Richardson, A. M., Julie, A., Cornes, K., Godwin, H. J., Garner, M., Simon, P., & Donnelly, N. (2018). Searching for two categories of target in dynamic visual displays impairs monitoring ability, (August 2017), 440–449. https://doi.org/10.1002/acp.3416

Riggs, C. A., Cornes, K., Godwin, H. J., Liversedge, S. P., Guest, R., & Donnelly, N. (2017). The importance of search strategy for finding targets in open terrain. *Cognitive Research: Principles and Implications*, *2*(1), 14. https://doi.org/10.1186/s41235-017-0049-4

Riggs, C. A., Godwin, H. J., Mann, C. M., Smith, S. J., Boardman, M., Liversedge, S. P., & Donnelly, N. (2018). Rummage search by expert dyads, novice dyads and novice individuals for objects hidden in houses. *Visual Cognition*, 1–17. https://doi.org/10.1080/13506285.2018.1445678

Schwaninger, A., Bolfing, A., Halbherr, T., Helman, S., Belyavin, A., & Hay, L. (2008). The impact of image based factors and training on threat detection performance in X-ray screening. In *3rd International Conference on Research in Air Transportation, ICRAT* (pp. 317–324). Fairfax, Virginia, USA.

Schwark, J. D., MacDonald, J., Sandry, J., & Dolgov, I. (2013). Prevalence-based decisions undermine visual search. *Visual Cognition*, *21*(5), 541–568. https://doi.org/10.1080/13506285.2013.811135

Smith, A. D., Wallace, F., Hood, B., & Gilchrist, I. D. (2009). Mechanisms of large-scale environmental search: probability cueing depends on the relationship between landmarks and target distribution. *Cognitive Processing*, *10 Suppl 2*, S305-6. https://doi.org/10.1007/s10339-009-0312-9

Sterchi, Y., Hättenschwiler, N., & Schwaninger, A. (2019). Detection measures for visual inspection of X-ray images of passenger baggage. *Attention, Perception, and Psychophysics*, *81*(5), 1297–1311. https://doi.org/10.3758/s13414-018-01654-8

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. https://doi.org/doi:10.1037/a0025887

Szczerba, J., & Hersberger, R. (2014). The Use of Stereoscopic Depth in an Automotive Instrument Display: Evaluating User-Performance in Visual Search and Change Detection. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 1184–1188. https://doi.org/10.1177/1541931214581247

van Beurden, M. H. P. H., IJsselsteijn, W. A., & Juola, J. F. (2012). Effectiveness of stereoscopic displays in medicine: A review. *3D Research*, *3*(1), 1–13.

https://doi.org/10.1007/3DRes.01(2012)3