Acta Psychologica 219 (2021) 103394

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Training detection of camouflaged targets in natural scenes: Backgrounds and targets both matter

Mark B. Neider [a,*], Dawn M. Sarno [b], Joanna E. Lewis [c], Ada D. Mishler [d], Alyssa S. Hess [e], Corey J. Bohil [a], Arthur F. Kramer [f,g]

[a] University of Central Florida, United States of America
[b] Clemson University, United States of America
[c] University of Northern Colorado, United States of America
[d] Naval Air Warfare Center Training Systems Division, United States of America
[e] Expedia Group, United States of America
[f] Northeastern University, United States of America
[g] University of Illinois, United States of America

ARTICLE INFO

Keywords:
Visual search
Training
Camouflage
Naturalistic scenes
Perceptual learning

ABSTRACT

As target-background similarity increases, search performance declines, but this pattern can be attenuated with training. In the present study we (1) characterized training and transfer effects in visual search for camouflaged targets in naturalistic scenes, (2) evaluated whether transfer effects are preserved 3 months after training, (3) tested the suitability of the perceptual learning hypothesis (i.e., using learned scene statistics to aid camouflaged target detection) for explaining camouflage search improvements over training, and (4) provide guidance for camouflage detection training in practice. Participants were assigned to one of three training groups: adaptive camouflage (difficulty varied by performance), massed camouflage (difficulty increased over time), or an active control (no camouflage), and trained over 14 sessions. Additional sessions measured transfer (immediately post training) and retention of training benefits (10 days and 3 months post training). Both the adaptive and massed training groups showed improved camouflaged target detection up to 3 months following training, relative to the control. These benefits were observed only with backgrounds and targets that were similar to those experienced during training and are broadly consistent with the perceptual learning hypothesis. In practice, training interventions should utilize stimuli similar to the operational environment in which detection is expected to occur.

We search for objects constantly, our keys in the kitchen, our car in the parking lot, our shirt in the closet. The traditional search paradigm is straightforward; an observer is asked to find a target amongst an array of distractors (see Wolfe, 1998, for a review). Target presence and set size (i.e., the number of items in the search array) are common manipulations. Search objects are most often viewed on a homogenous background and spaced randomly. Studies of this nature have taught us a great deal about the low-level visual features typically extracted by the visual system (Julesz, 1981; Treisman & Gelade, 1980; Treisman & Gormican, 1988) and the manner in which those features are used to guide attention toward likely target objects (Motter & Belky, 1998; Wolfe, 1994; Wolfe et al., 1989; Zelinsky, 1996).

Search in naturalistic scenes, however, is often more complex than what is encountered in many traditional search tasks (e.g., discrete items, such as Ts and Ls presented on a homogenous background). One factor that is often discounted in laboratory studies is the extent to which search targets and the backgrounds upon which they appear share common features. Wolfe et al. (2002) explored search behavior in several traditional laboratory tasks in which the search objects were presented on complex backgrounds. They found that increases in target-background similarity typically produced a corresponding increase in response times (RT); the more similar a target is to the background it is presented on, the more difficult it is to find. In follow-up work, Neider and Zelinsky (2006b) developed a paradigm to look more directly at visual search for camouflaged targets in which real world objects were used as the search items (e.g., children's toys). For each object a

* Corresponding author at: Department of Psychology, University of Central Florida, 4111 Pictor Lane, Psychology Bldg 99 Ste. 320, Orlando, FL 32816-1390, United States of America.

E-mail address: mark.neider@ucf.edu (M.B. Neider).

https://doi.org/10.1016/j.actpsy.2021.103394
Received 9 November 2020; Received in revised form 4 August 2021; Accepted 9 August 2021
Available online 12 August 2021

corresponding background was created by repeating a portion of the target item over a canvas. The background was then used as an underlay for the corresponding target object and additional distractor items, resulting in a search task in which the target appeared to blend into the background while the distractors remained salient. Traditional search measures (e.g., target presence, set size) were manipulated. Since the target and background in this paradigm share similar features, a good strategy for locating the target is to ignore salient distractor objects and restrict search to the background regions of the display. Observed RTs followed a pattern similar to those reported by Wolfe et al. (2002); however, supplemental eye movement analyses revealed that in the course of searching, observers preferred to make eye movements to salient target-dissimilar distractor items while target-similar background regions were relatively neglected (Neider & Zelinsky, 2006b).

It has been shown that observers trained on search images generated using Neider & Zelinsky's method demonstrate large performance improvements in both RTs and accuracy. When trained observers are asked after training to search for novel camouflaged targets amongst novel distractors, near perfect transfer is observed; trained observers exhibit no performance cost for novel stimuli (Boot et al., 2009; Neider et al., 2010). This observation of near perfect transfer in camouflage search is surprising considering the existing perceptual learning literature showing that transfer of training tends to be extremely limited (e.g., Ahissar & Hochstein, 1996; Ball & Sekuler, 1982). But what are the perceptual and cognitive mechanisms underlying skill acquisition and transfer of training in this context, and to what extent do such findings hold up when search displays are more representative of the environments we encounter in the real world, as opposed to highly structured arrays of items? Along these lines, Chen and Hegdé (2012) corroborated previously observed training improvements using an alternative method for creating camouflage displays that produced images more consistent, in terms of overall scene statistics, with what we encounter in the natural world. Although the search images were perhaps better described as natural textures than scenes (i.e., the search scenes were akin to extreme close-ups of natural images and thus lacked some information typically inherent to scenes; see Võ, 2021), their findings provide an important data point in our understanding of search for camouflaged targets and associated training. Critically, Chen and Hegdé (2012) found that learning to break camouflage and successfully locate a target object was more dependent on the background upon which the target appeared than on the target being searched for. They suggested that over time searchers become accustomed to the scene statistics associated with a given background, and then employ this perceptual learning of task-relevant image statistics to aid in low level visual processing, such as figure-ground assignments. Neider et al. (2013) provided additional evidence for a perceptual-tuning account of camouflage search training improvements using a rapid presentation search paradigm.

The previously described studies provide significant steps toward a better understanding of the processes underlying search for camouflaged targets. However, those studies also suffer from shortcomings that make it difficult to assess whether their findings can offer meaningful insights for improving camouflage search performance in more realistic settings. The first of these relates to the nature of training and retention. Specifically, the previously described studies that explored training effects exposed participants to relatively short training regimens and generally characterized training effects directly following training, leaving questions of longer-term preservation of training effects unanswered. The second shortcoming relates to the nature of the stimuli employed in these studies. In the case of studies done by Neider and colleagues (Boot et al., 2009; Neider et al., 2010; Neider & Zelinsky, 2006a), the stimulus set was composed of toy objects overlaid on a similar, but highly structured and artificial background. Although Chen and Hegdé (2012) did not use structured backgrounds in their work, their task, in which participants searched for a synthesized “digital embryo” target on a natural background (backgrounds were viewed extremely close up and no scene-like properties were discernible, thus the background itself was better described as a natural texture), was not representative of what an individual in the real world might encounter, where environments, and the objects that exist within them, not only have visual properties, but conceptual ones as well (e.g., object-spatial relationships; Neider & Zelinsky, 2006b). To fill this gap between lab-based search tasks and naturalistic search environments, we developed a paradigm that extends previous work on search for camouflaged targets to fully naturalistic scenes in which the search task can be presented in contexts that better reflect what might be encountered in a natural environment (Hess et al., 2016). Specifically, Hess and colleagues had participants search naturalistic scenes for camouflaged target patches that varied in size (40, 50, 60, and 70 pixels). Target patches were used rather than discrete objects based on the assertion that under realistic conditions, well camouflaged target objects would not be very object-like at all, and instead would be best characterized as a visual anomaly in a scene; an easily segmented object would represent a poor attempt at camouflage. Overall, RTs decreased as targets got larger (i.e., easier to detect) and accuracy was generally low at around 70% (see Neider & Zelinsky, 2006a, for a similar pattern of results). Based on the data, it is reasonable to infer that participants found the task difficult (see Kristjánsson, 2015, and Palmer et al., 2011, for examples of other challenging search tasks). However, despite the difficulty of the search task, and of relevance to the current study, participants displayed robust improvements in RT as they became more familiar with the task (block 4 compared to block 1), even at the very smallest target size. By block 4, performance for the hardest targets was nearly as good as for the easiest targets. Performance in terms of accuracy followed a similar pattern to RT.

The current study builds on previous work by conducting a large-scale training study in which participants learned to search for camouflaged targets in fully naturalistic scenes over the course of 14 training sessions. The goals of the study were to (1) extend previous studies on visual search training and transfer for camouflaged targets to naturalistic scenes, (2) characterize the extent to which any performance benefits attained through training are preserved after training, (3) directly test the robustness of the perceptual learning hypothesis in explaining camouflage search improvements associated with training, and (4) provide data to inform future development of camouflage detection training tasks and procedures that could be employed in real world contexts. To address goal one, we used targets and scenes that were similar to those used by Hess et al. (2016). These were images of natural forest and urban environments within which we embedded a camouflaged target patch. Participants responded on each trial whether a target was present or absent across multiple difficulty levels and across multiple experimental sessions. To address goal two, we brought participants back to the lab for follow-up testing approximately 10 days and 3 months after completion of their training program. The extent to which training and transfer benefits are preserved over time is important for informing real world training to any sort of search task in which the target signal is embedded in highly correlated noise, including both medical and military contexts. To address goal three, we had two transfer conditions. Participants were trained to search for a camouflaged target embedded in natural forest images. At transfer, participants performed the same task, but we varied the nature of the target and backgrounds on which the targets appeared. In varying background classes, participants performed the search task on novel backgrounds that were of the same class as the training images (forest images), as well as on backgrounds that were both perceptually and semantically different from those used in training (urban images). In varying target classes, participants searched for three novel target types that were produced in a manner similar, but not identical, to the targets during training, and then embedded on previously trained natural forest backgrounds (see Method section for additional details). If training and transfer benefits were strictly associated with improved perceptual learning of background statistics, then we would expect to find transfer of training to novel backgrounds that are similar in class to those used

Table 1
Timeline of study completion. Note that only sessions 4-21, 23, and 24 (italicized and bolded) are reported in the current analyses.

Sessions 1-2	Session 3	Sessions 4-5	Sessions 6-19	Sessions 20-21	Session 22	Session 23	Session 24
Pre-screening/ Cognitive Battery	Pre-training fNIRs/ Eye tracking (Transfer tasks)	***Pre-training (Transfer tasks)***	***Camouflage training***	***Post-training (Transfer tasks)***	Post-training fNIRs/Eye tracking (Transfer tasks)	***10-day Retention (Transfer tasks)***	***3-month Retention (Transfer tasks)***

Fig. 1. An illustration of the process by which target locations/sizes were captured from each background image.

Fig. 2. Example target at size 100.

Table 2
Mean counts for each target size for the training groups.

Target Size	100	90	80	70	60	50	40
Adaptive	164	188	222	223	208	185	133
Massed	192	192	192	192	192	192	192

Note. Values displayed represent the mean number of each target size (in pixels) presented on target present trials across participants and training session.

for training, and to novel target classes embedded in such backgrounds. In contrast, we would not expect to find transfer of training to search images associated with novel background classes. To address goal four, we had three different training groups. The first group (active control) spent their 14 training sessions engaged in a traditional search task (find a T amongst rotated Ls). The second group (massed training) searched for camouflaged targets in natural forest images. The difficulty of the search task was increased every two training sessions. The third group (adaptive training) was like the second; however, training difficulty was adapted to individual performance using a 3 up, 1 down staircase method. Previous research has indicated that adaptive training, in which task difficulty is determined by the participant's performance, results in the most robust training and transfer benefits (Von Bastian & Oberauer, 2014). Therefore, we expected that although both training groups would outperform the control group, the adaptive group would demonstrate more robust transfer than the massed group.

1. Method

1.1. Participants

A total of 90 (M_{age} = 20.21, 59 female) participants were recruited from the University of Central Florida and were compensated \$5 per hour with a bonus of an additional \$5 per hour if they completed the first 22 sessions of the study. All participants were prescreened for visual acuity (20/32 or better corrected vision on a Snellen eye chart), color vision (Ishihara's test for color blindness; 13 plates), and stereopsis (Stereo Fly test).

Sixty-one participants completed the first 22 sessions of the study (attrition rate approximately 30%). An additional 8 participants were removed due to below chance performance (50% accuracy) on more than half of the training sessions. Our final sample size was 53 participants for the main analyses, 40 participants for the 10-day retention analyses, and 23 participants for 3-month retention analyses.

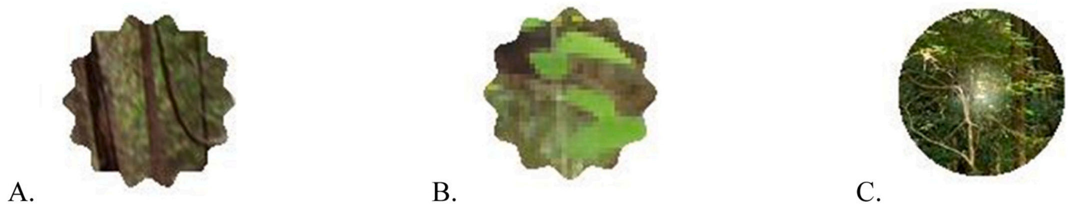

Fig. 3. Examples of novel target categories, blur (A), geometric (B) and lens flare (C).

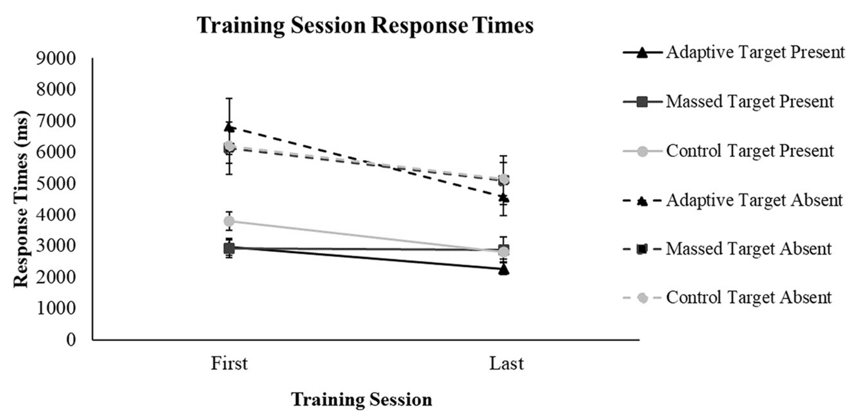

Fig. 4. Training session response times by training group.
Note. Error bars represent the standard error of the mean. Adaptive $N = 16$, Massed $N = 18$,

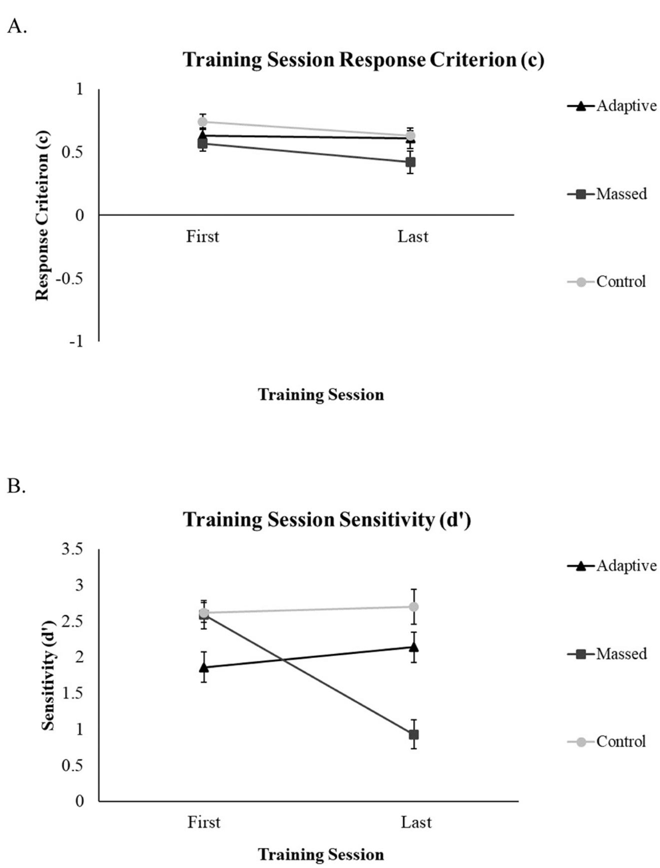

Fig. 5. Training session signal detection analyses by training group, response criterion (c) (A) and sensitivity (d') (B).
Note. Error bars represent the standard error of the mean. Adaptive $N = 16$, Massed $N = 18$,

1.2. *Study structure*

The entire study took place over 24 sessions ranging from an hour to 2 h in length (see Table 1 for a full outline of the sessions). The first two sessions consisted of a battery of cognitive and personality tests. In Session 1, participants were also screened for normal visual acuity, color vision, and depth perception, and then completed a brief demographic questionnaire. Afterwards, they completed a pencil-and-paper packet containing the trail making A and B tests (Reitan & Wolfson, 1985), Raven's Standard Progressive Matrices (Raven et al., 2000), letter and pattern comparison tests (Salthouse & Babcock, 1991), the letter sets test (Ekstrom et al., 1976), and the WAIS-III digit symbol coding and digit span subtests (Wechsler, 1997). Finally, participants completed computerized versions of the Ten Item Personality Inventory (Gosling et al., 2003), the Self-Determination Scale (Sheldon, 1995), and the Grit Scale (Duckworth et al., 2007). In Session 2, participants completed computerized versions of a spatial 2-back task, the Attention Network Test (Fan et al., 2002), a flanker task (Eriksen & Eriksen, 1974; Salthouse, 2010), and a Stroop task (Stroop, 1935). Functional near-infrared spectroscopy (fNIRS) data and oculomotor data were recorded during Sessions 3 and 22 while participants completed the camouflage task. Pre-training performance was assessed for transfer tasks in Sessions 4 and 5. Following training, performance was again assessed on transfer tasks in Sessions 20 and 21. Training took place over fourteen one-hour sessions. Retention of training was assessed 10 days and 3 months after Session 22. The results presented here focus on manual response time and error data from Sessions 4-21, 23 and 24. The cognitive battery and neuroimaging/eye tracking sessions were included to determine if various cognitive abilities influenced training abilities, and characterize the influence of training on cognitive and neural activity, respectively. However, these data are not reported here.

1.3. *Stimuli*

The stimuli for the search images were created by segmenting 30 images of natural forest scenes into 5×5 grids. For the training sessions,

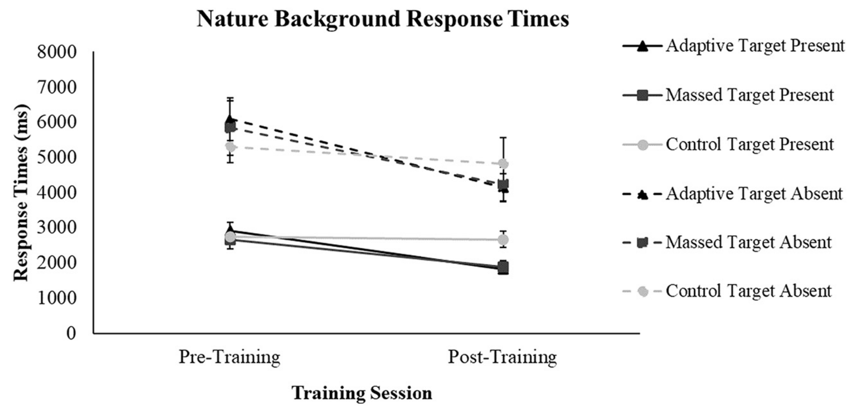

Fig. 6. Forest background response times by training group.
Note. Error bars represent the standard error of the mean. Adaptive $N = 17$, Massed $N = 18$, Control $N = 18$.

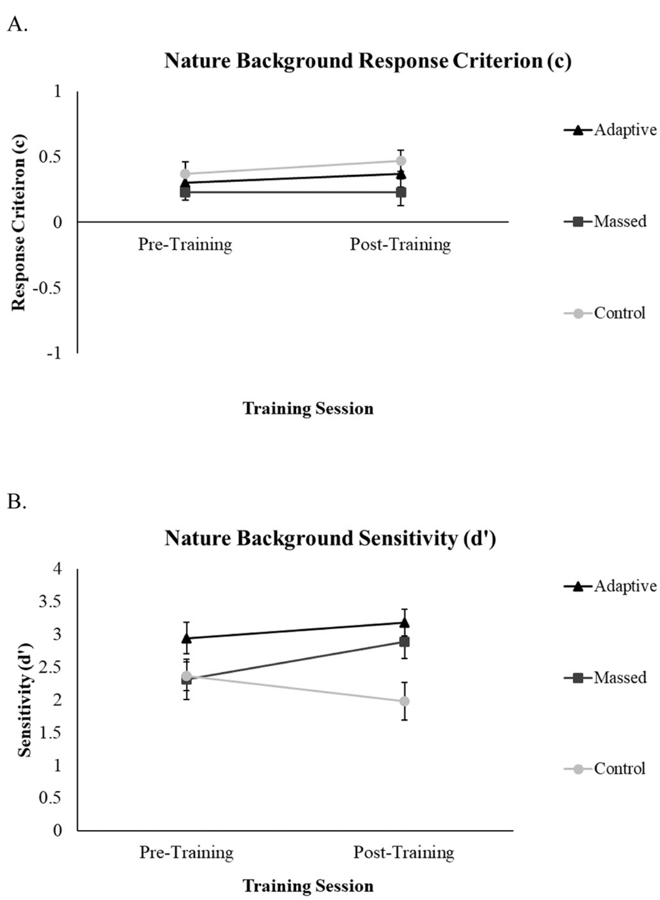

Fig. 7. Forest background signal detection analyses by training group, response criterion (c) (A) and sensitivity (d') (B).
Note. Error bars represent the standard error of the mean. Adaptive N = 17, Massed N = 18, Control N = 18.

the center point of each grid was used as an anchor point from which to create image patches of seven different sizes at each location (100 (2.7° x 2.8°), 90 (2.4° x 2.5°), 80 (2.0° x 2.0°), 70 (1.9° x 1.9°), 60 (1.6° x 1.7°), 50 (1.4° x 1.4°), and 40 (1.1° x 1.1°) pixels), that then served as targets for the corresponding image from which it was created. With the center point of the 5×5 grid excluded, this created 5040 unique targets, or 168 per image (see Fig. 1).

The camouflage target patch was created in Matlab by first selecting a circle patch of image pixels using a diameter equal to one of the seven sizes. Using a circular sine wave, which was scaled based on the patch size, we used this mask to select the target portion of the image RGB matrices. We generated an additional binary mask of 2 alternating sine waves, and then mapped the target matrices onto the mask to warp the patch image. This distortion changes the perspective, so we rotated the image back to the original position. Finally, we replaced the original patch with the distorted patch in the image matrices (see Fig. 1). The ultimate result of this process was a target that was extremely similar to, yet slightly different from, the underlying image region from which it was created. In other words, the target might be best described as a part of the image in which something is slightly amiss.

1.4. *Training groups*

The study included three training groups. All three training groups were trained over fourteen one-hour sessions. Both the massed and adaptive group received the same trial sequence. A preview of the isolated camouflage target patch appeared for 1000 milliseconds. A fixation cross at the center of the screen was then presented for 1000 milliseconds, followed by the search display. On half of the trials the camouflage target was present in the display (see Fig. 2 for an example) and on the other half of trials no target was present; presence was randomized across trials. Participants were asked to indicate, via button-press, if the target was present. Each session, regardless of training group, consisted of 192 trials broken down into four blocks of 48 trials, with optional breaks between blocks. Participants received feedback after each trial. The three training groups were:

1.4.1. *Adaptive*

Participants in the adaptive training group began each training session with the largest target size (i.e., 100 pixels). Their performance in the camouflaged task dictated the difficulty (i.e., target size) of the next trial based on a 3-up, 1-down staircase. Specifically, when participants completed three correct trials in a row, the difficulty of the task increased (i.e., the target size became smaller); when participants completed one trial incorrectly the difficulty of the task decreased (i.e., the target size became larger).

1.4.2. *Massed*

For participants who received massed training, the difficulty of the task was increased systematically over the course of the training sessions. Participants in this group were given two training sessions for each target size. Each participant started with the easiest and largest target size (i.e., 100), with the target sizes getting progressively smaller and more challenging across the course of training. See Table 2 for a comparison of the counts of each target size for the adaptive and massed training groups.

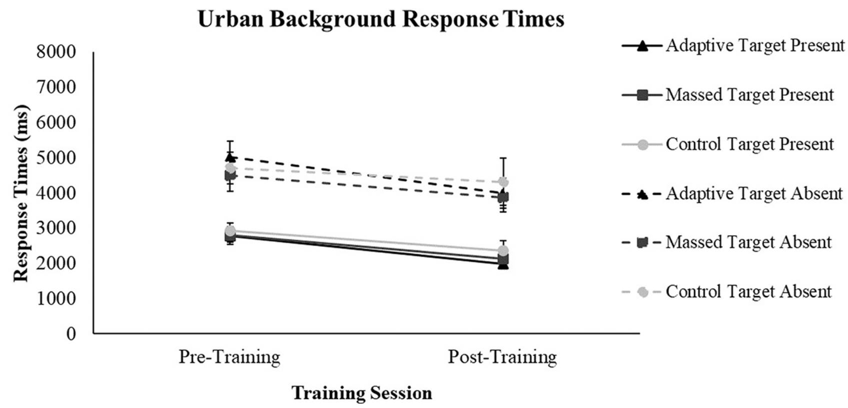

Fig. 8. Urban background response times by training group.
Note. Error bars represent the standard error of the mean. Adaptive $N = 17$, Massed $N = 18$, Control $N = 18$.

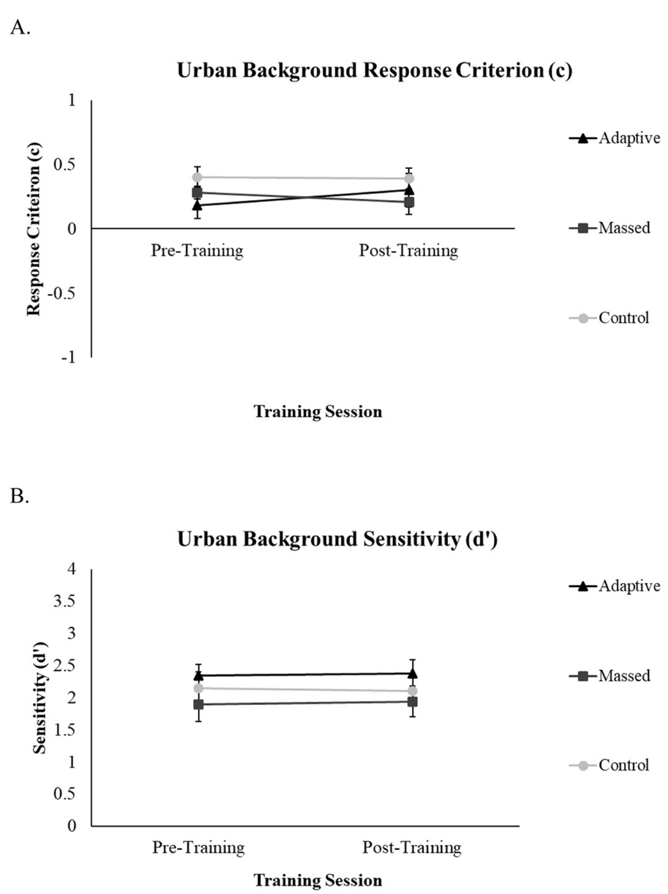

Fig. 9. Urban background signal detection analyses by training group, response criterion (c) (A) and sensitivity (d') (B).
Note. Error bars represent the standard error of the mean. Adaptive N = 17, Massed N = 18, Control N = 18.

1.4.3. Active control

Participants in the active control group received training on a non-camouflage task to serve as a baseline to compare to the training groups. In this condition, participants were presented with a simple visual search task, looking for a white T amongst white L's at various orientations on a black background. Similar to the massed training group, difficulty increased systematically over the course of training. Participants started with a set size of 40 and gradually increased to a set size of 100 over the course of training.

1.5. Transfer tasks

1.5.1. Novel backgrounds

Transfer to novel backgrounds involved untrained stimuli from the same category (forest images) or a different category (urban images) as training. The nature of the target was again specific to the background with which it was associated and the manner in which the target was created was identical to the process used during training (see Fig. 1). The procedure and stimuli were exactly the same as during training, with the following exceptions. Only target sizes 100 (see Fig. 2) and 40 were utilized in the transfer tasks. Each transfer task consisted of 160 trials broken up into four blocks of 40 trials with optional breaks between blocks. Novel backgrounds were manipulated within subjects; all participants searched both urban and forest images.

1.5.2. Novel targets

Transfer to novel target categories involved generating unique stimuli from three novel target classes within the same backgrounds (i.e., forest) as training. Three different image processing techniques were utilized to create three unique novel target categories: motion blur, geometric, and lens flare (see Fig. 3). These targets were intended to simulate targets in the environment such as someone moving (i.e., motion blur targets), a uniform (i.e., geometric targets), or something reflective (i.e., lens flare). Targets were created in Matlab. Motion blur and geometric targets both used a circular sine wave to define the outer edge of the target patch. Motion blur was created using a Matlab native image processing 2-D motion filter. We used linear motion by 4 pixels and 115 degrees. Geometric images were created using a block procedure for 3×3 pixel blocks which were averaged within each RGB matrix before finding the mean color across channels. The original 3×3 block was replaced with the new averaged color value. The lens flare patches were created using an array with incremental values based on the patch size to create 10 values. These values were used to lighten a circle within the target patch, which grew in radius with each iteration until reaching the target size, thus creating a diffused sense of light which was scaled in density based on the target size.

2. Results

For all analyses, we used SPSS 23 for Windows with default settings. Response times were computed based on correct trials. Signal detection theory (SDT) measures were included to determine the efficacy of the training groups (Green & Swets, 1988). Improvements in sensitivity (d') would indicate that our training resulted in improved ability to detect the camouflaged targets. Response criterion (c) was also analyzed to determine if the training changed participant's bias toward indicating if a target was present or absent. Lenient responders have scores that are

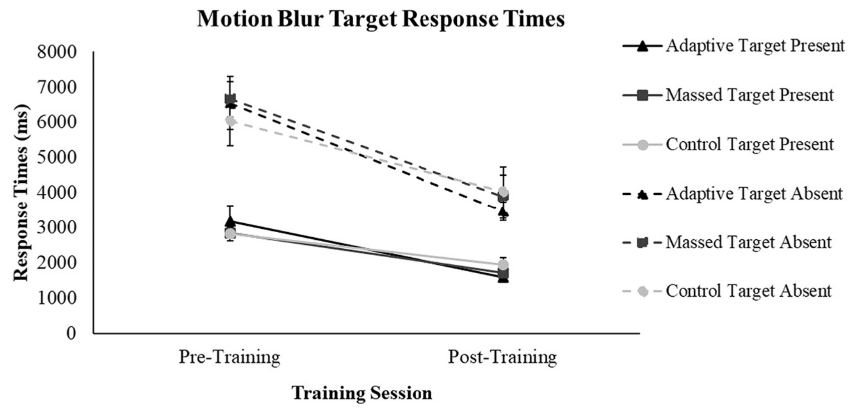

Fig. 10. Motion blur target response times by training group.
Note. Error bars represent the standard error of the mean. Adaptive $N = 17$, Massed $N = 18$, Control $N = 18$.

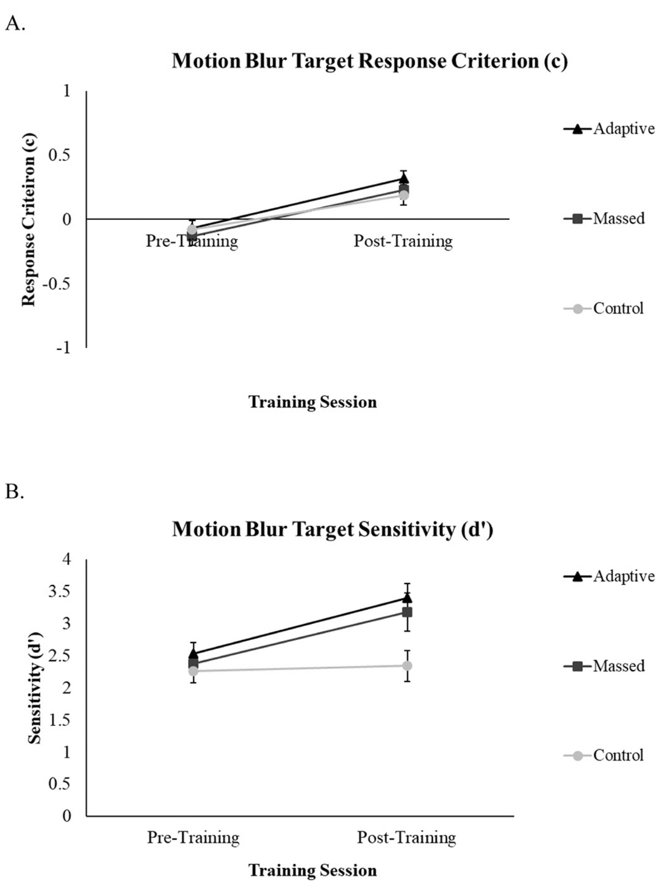

Fig. 11. Motion blur target signal detection analyses by training group, response criterion (c) (A) and sensitivity (d') (B).
Note. Error bars represent the standard error of the mean. Adaptive N = 17, Massed N = 18, Control N = 18.

less than zero and are more likely to say a target is present. Conservative responders have scores that are greater than zero and are more likely to say a target is absent. Unbiased responders have scores that are not statistically different from zero and are equally likely to say a target is present or absent (Stanislaw & Todorov, 1999). Additionally, to better characterize the nature of our effects, and particularly to inform cases in which the null was accepted, we analyzed posterior probabilities [$p_{BIC}(H1|D)$]. Posterior probabilities indicate a graded probability of whether the null hypothesis or alternative hypothesis is better supported based on the current data. A $p_{BIC}(H1|D) < 0.50$ suggests more evidence for the null hypothesis, whereas a $p_{BIC}(H1|D) > 0.50$ indicates more evidence for the alternative hypothesis (Masson, 2011).

2.1. *Training sessions*

To assess the effect of training, performance across training was analyzed separately for each training group. For each training group, we conducted a 2 (training session; first vs last session) x 2 (target presence; present vs absent) within-subjects factorial ANOVA on RT data. Additional one-way within-subjects ANOVAs with training session (first vs last) were conducted on each group's SDT data. We took this approach because each training group received different experiences during training, in terms of both the types of trials and the frequency of each difficulty level that they received, making session-by-session comparisons across groups impossible. For example, we would expect performance in the adaptive group to improve with training because they could experience all target difficulty levels within any given session, but the massed group's performance might decline over training because they were receiving harder search trials progressively across the training sessions. It should be noted that changes in behavior during training are not the critical measure for the current study. Instead, the primary focus was on pre- and post -training performance associated with the training groups, for which session-by-session comparisons across groups were possible and are included in the appropriate sections. These comparisons allowed us to determine which training approaches demonstrated the largest improvements in camouflaged target detection.

2.1.1. *Massed*

2.1.1.1. Response times. Massed participants' response times did not change over the course of training, $F(1,17) = 0.82$, $p = .379$, $\eta_p^2 = 0.05$, $p_{BIC}(H1|D) = 0.27$ (see Fig. 4). Massed participants were also slower on non-target trials $F(1,17) = 32.78$, $p < .001$, $\eta_p^2 = 0.66$, $p_{BIC}(H1|D) = 0.60$. There were no other significant effects (p's > 0.091). Given that the training task became more difficult for the massed training group as training progressed, the absence of a performance decline over sessions may be indicative of search processes becoming more efficient at extracting camouflaged targets over training.

Control $N = 18$.

2.1.1.2. Signal detection measures. Massed participants were less sensitive at the end training, $F(1,17) = 37.18$, $p < .001$, $\eta_p^2 = 0.69$, $p_{BIC}(H1|D) = 0.99$ (see Fig. 5B). These findings make sense intuitively; the massed group saw more challenging targets at the end of training, thus their ability to detect the target decreased. There was no effect of training session on response criterion, $F(1,17) = 2.13$, $p = .163$, $\eta_p^2 = 0.11$, $p_{BIC}(H1|D) = 0.03$ (see Fig. 5A).

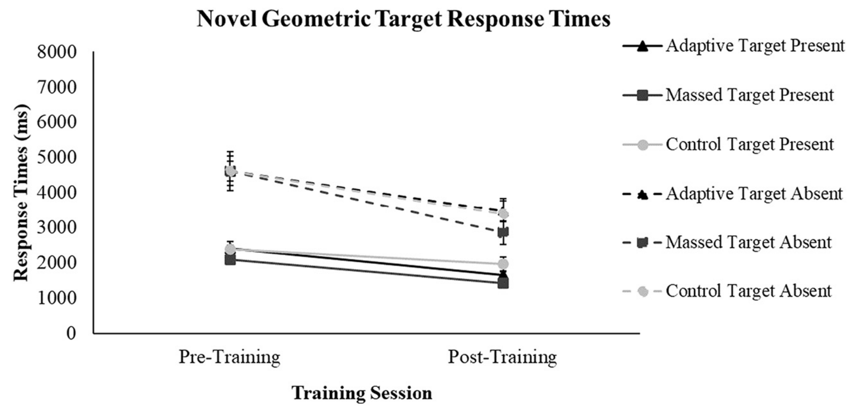

Fig. 12. Geometric target response times by training group.
Note. Error bars represent the standard error of the mean. Adaptive $N = 17$, Massed $N = 18$, Control $N = 18$.

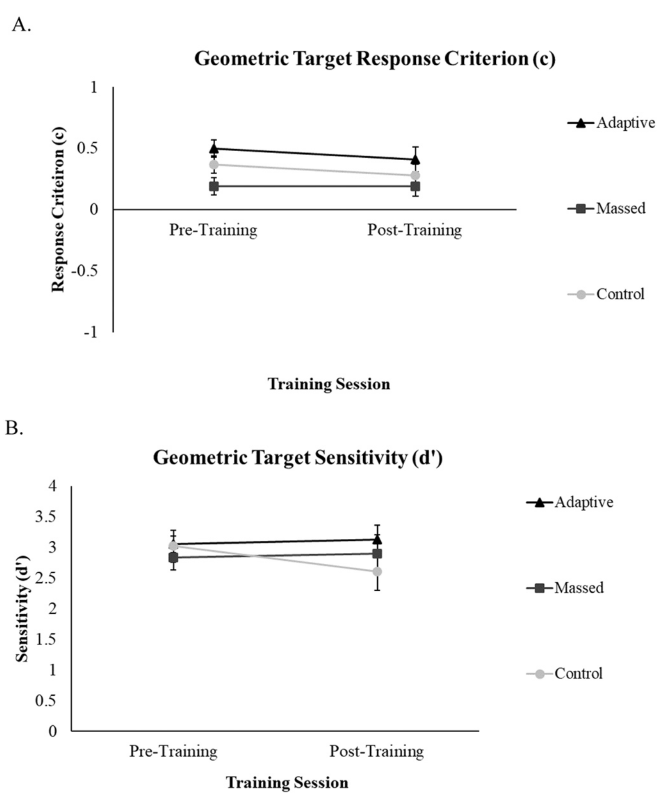

Fig. 13. Geometric target signal detection analyses by training group, response criterion (c) (A) and sensitivity (d') (B).
Note. Error bars represent the standard error of the mean. Adaptive $N = 17$, Massed $N = 18$, Control $N = 18$.

Control $N = 18$.

2.1.2. Adaptive

2.1.2.1. Response times. Adaptive participants responded more quickly at the end of training, $F(1,15) = 21.24$, $p < .001$, $\eta_p^2 = 0.59$, $p_{BIC}(H1|D) = 0.99$ (see Fig. 4), and on trials in which there was a target present, $F(1,15) = 33.63$, $p < .001$, $\eta_p^2 = 0.69$, $p_{BIC}(H1|D) = 0.99$. There was also an interaction of session and target presence, $F(1,15) = 21.95$, $p < .001$, $\eta_p^2 = 0.59$, $p_{BIC}(H1|D) = 0.99$. Post hoc tests revealed the influence of target presence was larger at the first session, $F(1,15) = 36.14$, $p < .001$, $\eta_p^2 = 0.71$, $p_{BIC}(H1|D) = 0.99$, than at the last session $F(1,15) = 25.64$, $p < .001$, $\eta_p^2 = 0.63$, $p_{BIC}(H1|D) = 0.99$.

2.1.2.2. Signal detection measures. Training did not influence the adaptive groups' sensitivity, $F(1,15) = 0.95$, $p = .345$, $\eta_p^2 = 0.06$, $p_{BIC}(H1|D) = 0.29$, or response criterion, $F(1,15) = 0.03$, $p = .856$, $\eta_p^2 < 0.01$, $p_{BIC}(H1|D) = 0.20$ (see Fig. 5).

Combined, these data indicate that participants in the adaptive training group became faster at locating camouflaged targets over the course of training without sacrificing their ability to detect the target.

2.1.3. Control

2.1.3.1. Response times. Control participants were faster over the course of training, $F(1,17) = 14.89$, $p = .001$, $\eta_p^2 = 0.47$, $p_{BIC}(H1|D) = 0.99$, and on trials in which there was a target present, $F(1,17) = 68.15$, $p < .001$, $\eta_p^2 = 0.80$, $p_{BIC}(H1|D) = 0.99$ (see Fig. 4). There were no other significant effects (p's > 0.808).

2.1.3.2. Signal detection measures. Training did not influence the control groups' sensitivity, $F(1,17) = 0.63$, $p = .805$, $\eta_p^2 < 0.01$, $p_{BIC}(H1|D) = 0.19$, nor their response criterion, $F(1,17) = 2.94$, $p = .105$, $\eta_p^2 = 0.15$, $p_{BIC}(H1|D) = 0.50$ (see Fig. 5).

Overall, these data suggest that, similar to the adaptive group, the control group became faster at the task over time, without decreasing in their ability to successfully differentiate between when the target was present versus when it was absent.

2.2. Transfer

Performance on the transfer tasks was analyzed to evaluate which training method engendered the best performance and to investigate the perceptual learning hypothesis. For each transfer task, we conducted 3 × 2 mixed factorial ANOVAs to examine differences between the groups in training outcomes for both novel backgrounds and novel target classes. Note that because performance for target size 40 was at or below chance for all participants, it was omitted from all analyses. Instead, we focus on performance at target size 100. Because search for a target that is present and one that is absent is qualitatively different (the former includes target guidance whereas the latter does not), here we conducted separate ANOVAs for target present and target absent trials, respectively. In all analyses, training group (adaptive vs massed vs control) was treated as a between-subjects variable, and transfer session (pre-training vs post-training) was treated as a within-subjects variable.

2.2.1. Novel backgrounds

2.2.1.1. Forest backgrounds

2.2.1.1.1. Target present response times. Overall, the training groups took roughly the same amount of time to find the target, $F(2,50) = 1.89$,

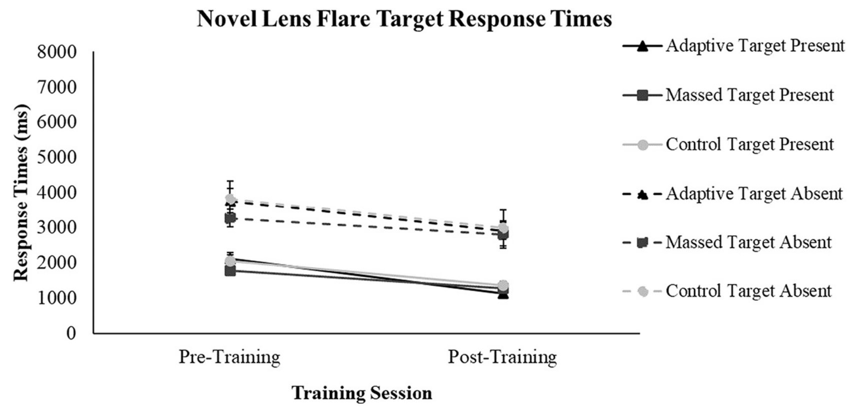

Fig. 14. Lens Flare target response times by training group.
Note. Error bars represent the standard error of the mean. Adaptive $N = 17$, Massed $N = 18$, Control $N = 18$.

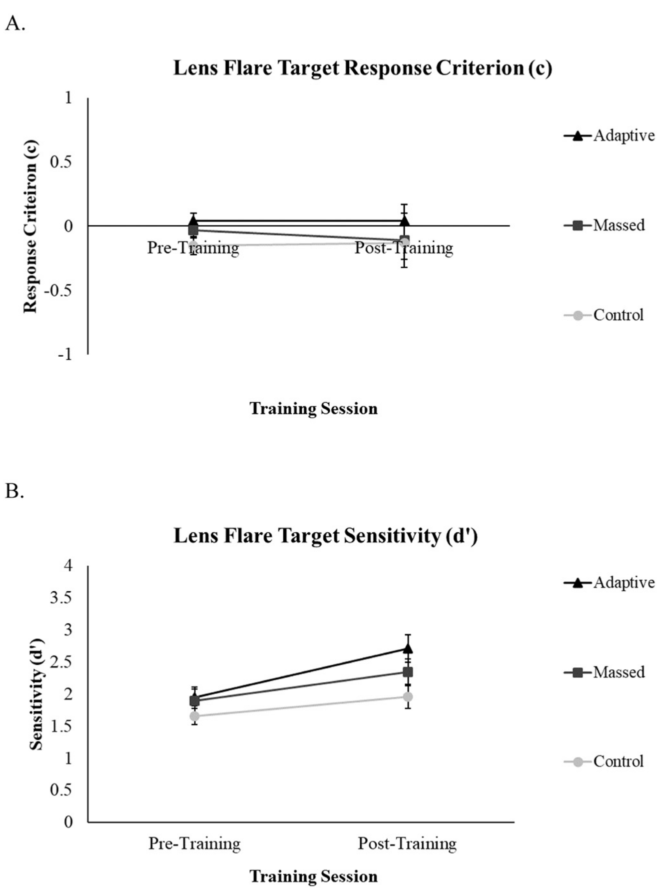

Fig. 15. Lens flare target signal detection analyses by training group, response criterion (c) (A) and sensitivity (d') (B).
Note. Error bars represent the standard error of the mean. Adaptive N = 17, Massed N = 18, Control N = 18.

$p = .162$, $\eta_p^2 = 0.07$, $p_{BIC}(H1|D) = 0.11$, (see Fig. 6). All participants became faster at the task over training, $F(1,50) = 22.40$, $p < .001$, $\eta_p^2 = 0.31$, $p_{BIC}(H1|D) = 0.99$, and there was a significant interaction between transfer session and training group, $F(2,50) = 4.86$, $p = .010$, $\eta_p^2 = 0.16$, $p_{BIC}(H1|D) = 0.66$. Analysis of simple effects indicated that prior to training, the groups did not differ in response times, $F(2,50) = 0.34$, $p = .716$, $\eta_p^2 = 0.01$, $p_{BIC}(H1|D) = 0.02$. However, group response times differed significantly after training, $F(2,50) = 6.41$, $p = .003$, $\eta_p^2 = 0.20$, $p_{BIC}(H1|D) = 0.87$. Bonferroni-corrected post hoc tests revealed that participants in the adaptive and massed groups were significantly faster than participants in the control group following training ($p = .007$, $p = .014$, respectively), but did not differ from each other ($p = 1.00$).

2.2.1.1.2. Target absent response times. All training groups took roughly the same amount time to respond on trials in which there was no target, $F(2,50) = 0.01$, $p = .993$, $\eta_p^2 < 0.01$, $p_{BIC}(H1|D) = 0.02$ (see Fig. 6), and all participants were faster at the task following training, $F(1,50) = 21.76$, $p < .001$, $\eta_p^2 = 0.30$, $p_{BIC}(H1|D) = 0.99$. There were no other significant effects (p's > 0.105).

2.2.1.1.3. Response criterion. There were no effects of training group, session, or their interaction on response criterion (p's > 0.160) (see Fig. 7A).

2.2.1.1.4. Sensitivity. Overall, participants were more sensitive in the task after training, $F(2,50) = 3.66$, $p = .033$, $\eta_p^2 = 0.13$, $p_{BIC}(H1|D) = 0.43$ (see Fig. 7B). There was also a significant interaction of transfer session and training group, $F(2,50) = 4.43$, $p = .017$, $\eta_p^2 = 0.15$, $p_{BIC}(H1|D) = 0.58$. Analysis of the simple effects of each training group for each session indicated that there was no difference between the groups for sensitivity prior to training. $F(2,50) = 1.75$, $p = .184$, $\eta_p^2 = 0.01$, $p_{BIC}(H1|D) = 0.02$. However, the groups differed after training, $F(2,50) = 6.01$, $p = .005$, $\eta_p^2 = 0.19$, $p_{BIC}(H1|D) = 0.83$. Bonferroni-corrected post hoc tests indicated that the adaptive and massed groups were more sensitive to the targets than the control group following training ($p = .005$, $p = .042$, respectively) but there was no difference between the adaptive and massed groups ($p = 1.00$) or any other significant effects (p's > 0.275).

Overall, participants in our training groups were able to detect the camouflaged targets in novel forest images both more quickly and more accurately than participants in our active control group. The active control group did perform faster following training, but demonstrated decreased sensitivity, perhaps reflecting a speed-accuracy tradeoff.

2.2.1.2. Urban backgrounds

2.2.1.2.1. Target present response times. All participants were faster after training for urban images, $F(1,50) = 17.96$, $p < .001$, $\eta_p^2 = 0.26$, $p_{BIC}(H1|D) = 0.26$ (see Fig. 8). However, there were no other significant effects (p's > 0.608).

2.2.1.2.2. Target absent response times. All participants were faster at the task following training for the target absent trials, $F(1,50) = 7.72$, $p = .008$, $\eta_p^2 = 0.13$, $p_{BIC}(H1|D) = 0.85$. There were no other significant effects (p's > 0.573).

2.2.1.2.3. Response criterion. There were also no significant effects on response criterion (p's > 0.250), indicating that training did not impact participants' response criterion (see Fig. 9A).

2.2.1.2.4. Sensitivity. Similar to the response criterion results, there were no significant effects on sensitivity (p's > 0.317), suggesting that training did not influence the participants' ability to detect the targets in

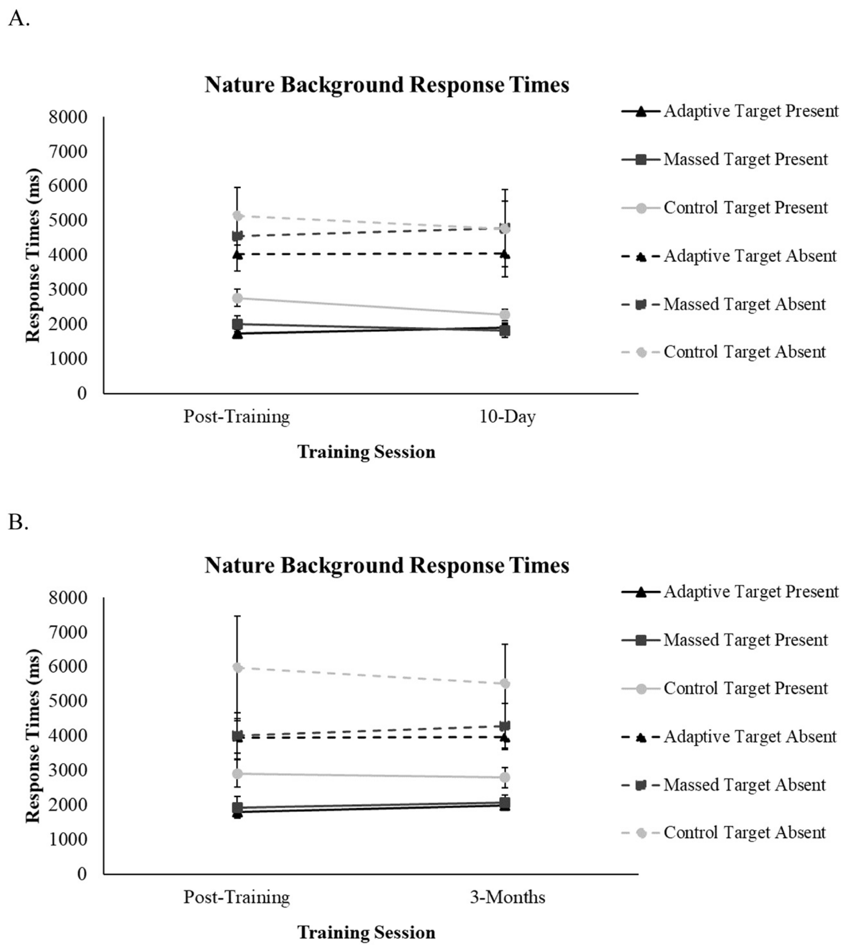

Fig. 16. Forest background 10-Day retention (A) and 3-Months (B) response times by training group.
Note. Error bars represent the standard error of the mean. 10-Day: Adaptive $N = 12$, Massed $N = 13$, Control $N = 15$. 3-Month: Adaptive $N = 8$, Massed N = 8, Control $N = 7$.

the urban images (see Fig. 9B).

2.2.1.2.5. Summary of novel background results. Overall, there were no differences between our two training groups and the active control for novel urban images. All groups did get significantly faster at the task but did not improve in their ability to detect the urban targets. The lack of transfer benefits for urban images, but strong transfer effects for novel forest images, is supportive of the perceptual learning hypothesis. Consistent with this account, these data suggest that after repeated exposures to forest images during training our participants may have become familiar with the underlying scene statistics associated with such scenes and then applied them to novel scenes of the same class at transfer (Chen & Hegdé, 2012; Neider et al., 2013).

2.2.2. Novel target classes

2.2.2.1. Motion blur target

2.2.2.1.1. Target present response times. All participants became significantly faster at detecting the motion blur target following training, $F(1,50) = 54.46$, $p < .001$, $\eta_p^2 = 0.52$, $p_{BIC}(H1|D) = 0.99$ (see Fig. 10). There were no other significant effects (p's > 0.203).

2.2.2.1.2. Target absent response times. Similar to target present trials, all participants became significantly faster at detecting the motion blur target following training, $F(1,50) = 59.50$, $p < .001$, $\eta_p^2 = 0.54$, $p_{BIC}(H1|D) = 0.99$ (see Fig. 10). There were no other significant effects (p's > 0.410).

2.2.2.1.3. Response criterion. All participants were more conservative after training, and classified fewer trials as target present, $F(1,50) = 31.27$, $p < .001$, $\eta_p^2 = 0.39$, $p_{BIC}(H1|D) = 0.99$ (see Fig. 11A) but there were no other significant effects (p's > 0.479).

2.2.2.1.4. Sensitivity. All participants were more sensitive toward the motion blur targets after training (see Fig. 11B), $F(1,50) = 17.50$, $p < .001$, $\eta_p^2 = 0.26$, $p_{BIC}(H1|D) = 0.99$. Training groups also differed in their sensitivities, $F(2,50) = 3.52$, $p = .037$, $\eta_p^2 = 0.12$, $p_{BIC}(H1|D) = 0.36$. Bonferroni-corrected post hoc tests revealed that the adaptive group was more sensitive than the control group ($p = .040$). These results should be cautiously interpreted given the low posterior probability of the main effect of group. There was also a significant interaction between transfer session and training condition, $F(2,50) = 3.52$, $p = .037$, $\eta_p^2 = 0.12$, $p_{BIC}(H1|D) = 0.36$. However, when analyzing the simple effects of each training group at each session individually, there was no difference between the groups for the pre-training session, $F(2,50) = 0.25$, $p = .778$, $\eta_p^2 = 0.10$, $p_{BIC}(H1|D) = 0.24$, or the post-training session, $F(2,50) = 0.78$, $p = .463$, $\eta_p^2 = 0.03$, $p_{BIC}(H1|D) = 0.04$. These findings indicate that the adaptive group may have been better than the control group at finding the motion blur targets, but this difference was probably not a result of training.

2.2.2.1.5. Summary of motion blur results. Overall, all participants demonstrated performance benefits for the motion blur targets. However, improvements in both sensitivity and response times were not specific to a training condition; camouflaged training groups were no better at locating the motion blur target after training than the control group, indicating that improvements were attributable to a test-retest benefit.

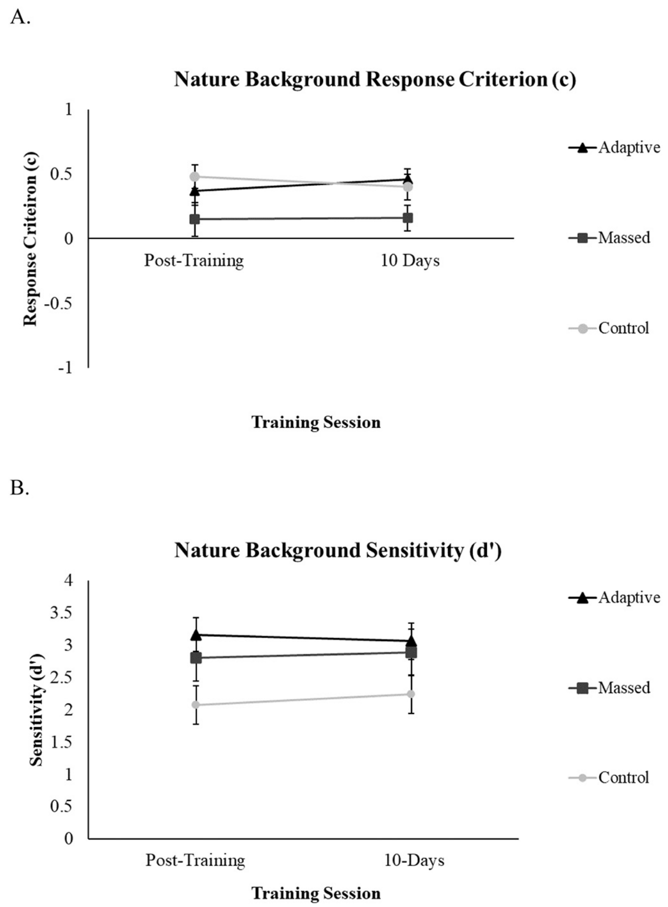

Fig. 17. Forest background 10-Day signal detection analyses by training group, response criterion (c) (A) and sensitivity (d') (B).
Note. Error bars represent the standard error of the mean. Adaptive $N = 12$, Massed $N = 13$, Control $N = 15$.

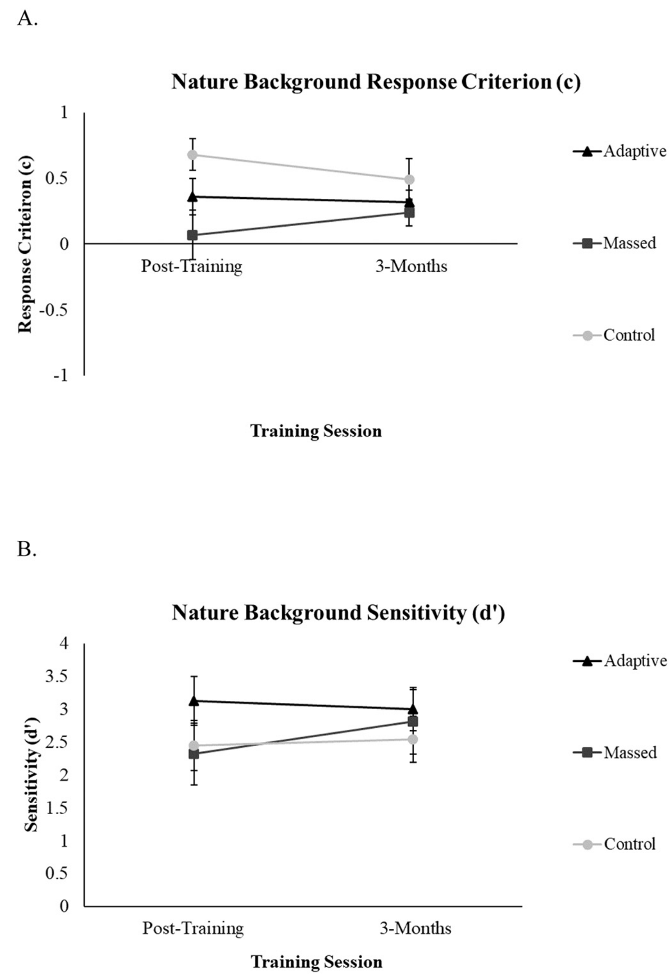

Fig. 18. Forest background 3-Months signal detection analyses by training group, response criterion (c) (A) and sensitivity (d') (B).
Note. Error bars represent the standard error of the mean. Adaptive $N = 8$, Massed N = 8, Control $N = 7$.

2.2.2.2. Geometric target

2.2.2.2.1. Target present response times. As with the motion blur targets, all participants became significantly faster at detecting the geometric target following training, $F(1,50) = 46.37$, $p < .001$, $\eta_p^2 = 0.48$, $p_{BIC}(H1|D) = 0.99$ (see Fig. 12). The main effect of training group did not quite reach significance, $F(2,50) = 3.15$, $p = .052$, $\eta_p^2 = 0.11$, $p_{BIC}(H1|D) = 0.29$. A Bonferroni-corrected post hoc test revealed that this trend was driven by the massed group being not quite significantly faster than the control group ($p = .052$). However, the training groups did not systematically vary across training, $F(2,50) = 1.29$, $p = .285$, $\eta_p^2 = 0.05$, $p_{BIC}(H1|D) = 0.07$, suggesting that any response time differences between the massed and control groups were present prior to the experiment and not a result of the training methodology.

2.2.2.2.2. Target absent response times. Again, all participants demonstrated faster response times across training, $F(1,50) = 41.81$, $p < .001$, $\eta_p^2 = 0.46$, $p_{BIC}(H1|D) = 0.99$ (see Fig. 12). There were no other significant effects (p's > 0.517).

2.2.2.2.3. Response criterion. The training groups differed in their response criterions, $F(2,50) = 4.85$, $p = .012$, $\eta_p^2 = 0.16$, $p_{BIC}(H1|D) = 0.66$ (see Fig. 13A). Bonferroni-corrected post hoc tests revealed that the adaptive group was more conservative than the massed group ($p = .009$). There were no other significant differences between the groups or other effects (p's > 0.339).

2.2.2.2.4. Sensitivity. There were no significant effects on sensitivity (p's > 0.301) (see Fig. 13B), suggesting that all participants were similarly able to discriminate target present/absent trials, regardless of training group or session.

2.2.2.2.5. Summary of geometric results. Overall, the geometric target results suggest that our training approaches did not transfer to this novel target category. All three training groups did demonstrate faster response times across training, but, as with the motion blur targets, this was likely associated with test-retest improvements.

2.2.2.3. Lens flare target

2.2.2.3.1. Target present response times. Similar to the first two novel target classes, all participants became faster at detecting the lens flare target over the course of training, $F(1,50) = 64.07$, $p < .001$, $\eta_p^2 = 0.56$, $p_{BIC}(H1|D) = 0.99$ (see Fig. 14), but there were no other significant differences (p's > 0.105).

2.2.2.3.2. Target absent response times. Again, participants demonstrated faster response times after training for target absent trials, $F(1,50) = 16.34$, $p < .001$, $\eta_p^2 = 0.25$, $p_{BIC}(H1|D) = 0.99$ (see Fig. 14). However, similar to the target present trials, there were no other significant differences (p's > 0.592).

2.2.2.3.3. Response criterion. There were no significant effects on response criterion ($ps > 0.074$) (see Fig. 15A).

2.2.2.3.4. Sensitivity. All participants were more sensitive after training, $F(1,50) = 23.25$, $p < .001$, $\eta_p^2 = 0.32$, $p_{BIC}(H1|D) = 0.99$, but there were no other significant effects (p's > 0.072) (see Fig. 15B).

2.2.2.3.5. Summary of lens flare results. Overall, all participants appeared to become more sensitive and faster over the course of training for the lens flare targets. However, these improvements were not dependent on which training group the participants were assigned to.

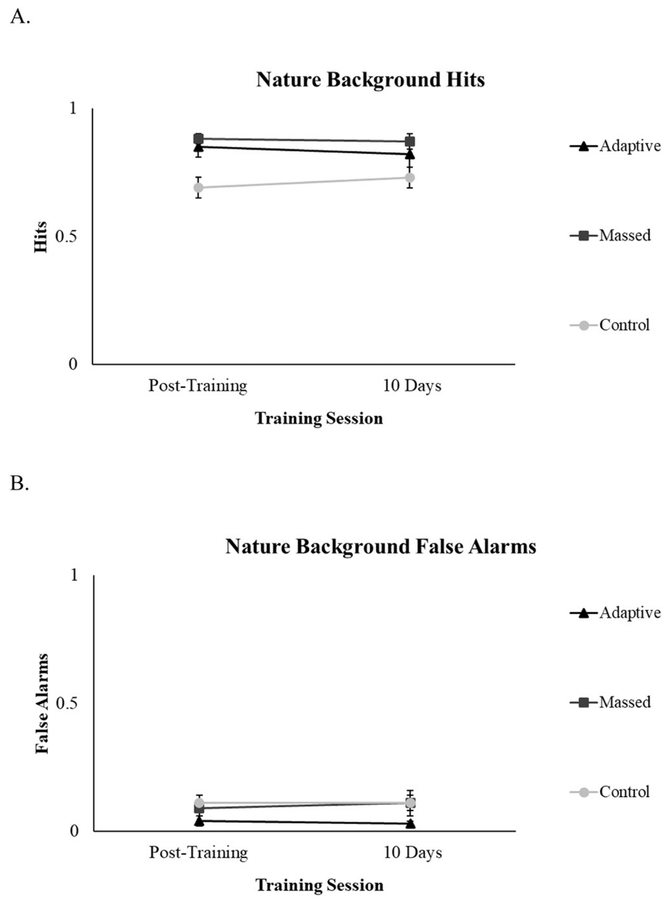

Fig. 19. Forest background 10-Day signal detection analyses by training group, hits (A) and false alarms (B).
Note. Error bars represent the standard error of the mean. Adaptive $N = 12$, Massed $N = 13$, Control $N = 15$.

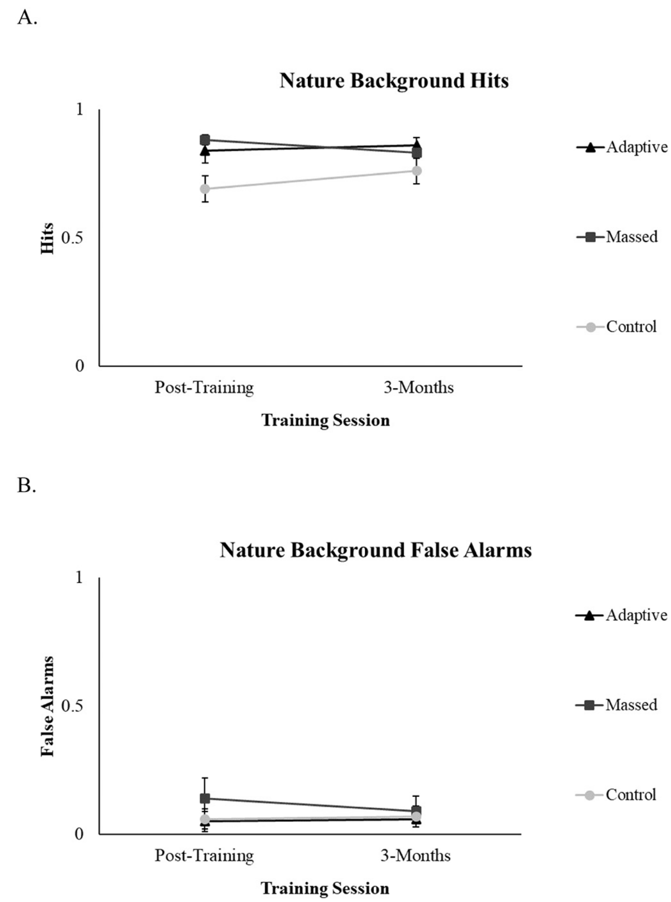

Fig. 20. Forest background 3-Months signal detection analyses by training group, hits (A) and false alarms (B).
Note. Error bars represent the standard error of the mean. Adaptive $N = 8$, Massed N = 8, Control $N = 7$.

Additionally, our training groups did not appear to benefit relative to the control group across all novel target classes. Broadly, as with the other two target categories, these findings suggest that our training intervention did not engender differential improvement in search for lens flare targets.

2.2.3. *Retention of training of novel backgrounds*

To evaluate the extent to which each training group retained any performance benefits (i.e., goal 2) associated with training, we conducted the same analyses as those for the transfer sessions, but compared performance from sessions 20 to 23, and 20 to 24, rather than session 5. As previously articulated, our participant pool experienced attrition at the 10-days and 3-month retention sessions; however, we had roughly the same number of participants in each group at both retention intervals (10-days: Adaptive: 12, Control: 15, Massed: 13; 3-months: Adaptive: 8, Control: 7, Massed: 8). Lastly, due to this attrition across retention sessions and the resulting decreased power/null effects, hits and false alarms were also analyzed to better categorize training retention. Hits represented the trials where a target was present, and participants correctly identified the target. False alarms, on the other hand, occurred on trials where there was no target present, but the participant indicated there was a target present. These additional measures may be more sensitive to capture the retention of any training benefits.

2.2.3.1. *Forest backgrounds*

2.2.3.1.1. Target present response times. Overall, participants were faster at the task 10 days after training compared to immediately following training, $F(1,37) = 4.41$, $p = .043$, $\eta_p^2 = 0.11$, $p_{BIC}(H1|D) = 0.62$. There were also significant effects of training group, $F(2,37) = 4.73$, $p = .015$, $\eta_p^2 = 0.20$, $p_{BIC}(H1|D) = 0.68$, and a significant main interaction of training group and retention session, $F(2,37) = 4.71$, $p = .015$, $\eta_p^2 = 0.20$, $p_{BIC}(H1|D) = 0.68$ (see Fig. 16B). However, simple effects revealed that there were no differences in response times between the groups 10 days later, $F(2,37) = 2.19$, $p = .126$, $\eta_p^2 = 0.11$, $p_{BIC}(H1|D) = 0.20$. This indicates any meaningful differences between the groups occurred immediately following training. Interestingly, there was a main effect of training group when comparing performance immediately after training and 3 months later, $F(2,20) = 5.18$, $p = .015$, $\eta_p^2 = 0.34$, $p_{BIC}(H1|D) = 0.76$, with the adaptive and massed groups being significantly faster than the control group (LSD post hoc tests; $p = .007$, $p = .016$, respectively) (see Fig. 16B). There were no other significant effects (p's > 0.605). Overall, these response time results suggest that 10 days later the meaningful differences disappear, but 3 months later benefits for our two training groups reemerged.

2.2.3.1.2. Target absent response times. In a similar vein to the initial transfer data, there were no significant effects on target absent response times for the retention data (p's > 0.256) (see Fig. 16). All groups appear to be roughly equivalent in target absent response times throughout retention.

2.2.3.1.3. Response criterion. There was a significant effect of training group on response criterion when comparing response criterion immediately post training to response criterion 10 days post training (see Fig. 17A), $F(2,37) = 3.41$, $p = .044$, $\eta_p^2 = 0.16$, $p_{BIC}(H1|D) = 0.45$. LSD post hoc tests revealed that the massed group was more liberal than the adaptive and control groups ($p = .044$, $p = .021$, respectively). There

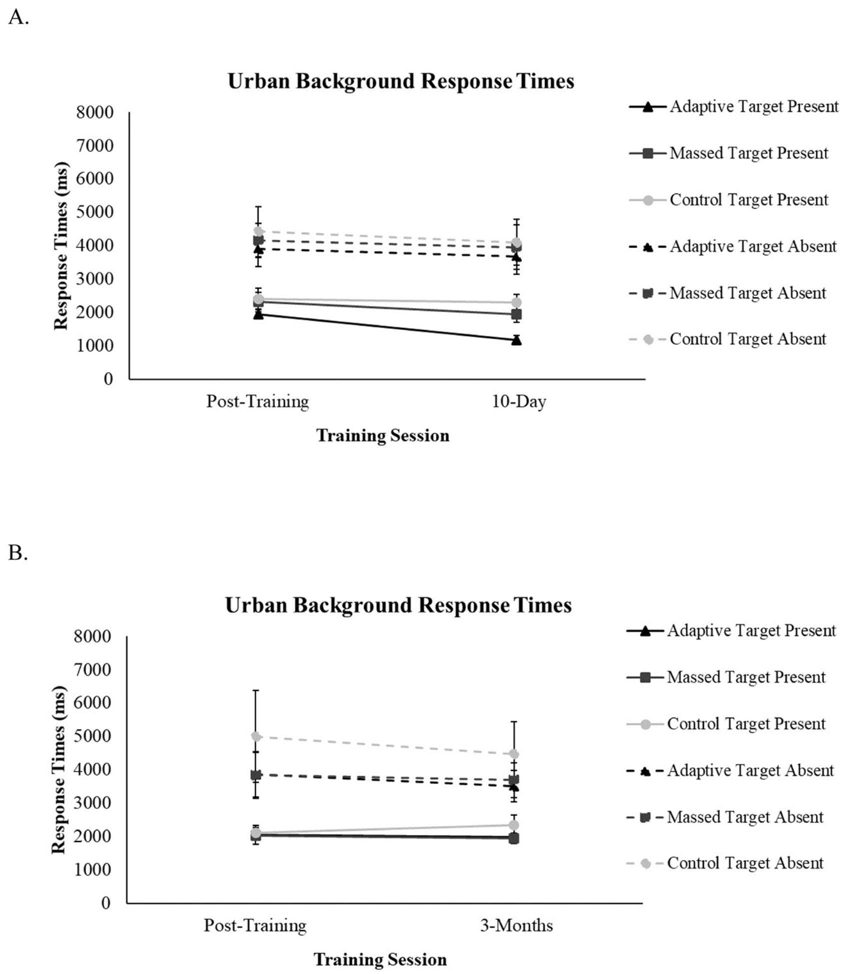

Fig. 21. Urban background 10-Day retention (A) and 3-Months (B) response times by training group.
Note. Error bars represent the standard error of the mean. 10-Day: Adaptive $N = 12$, Massed $N = 13$, Control $N = 15$. 3-Month: Adaptive N = 8, Massed N = 8, Control N = 7.

was no difference between the control group and the adaptive group ($p = .840$) and no other significant effects (p's > 0.573).

Similarly, when comparing the post training session to 3-months following training there was a main effect of training group, $F(2,20) = 3.89$, $p = .037$, $\eta_p^2 = 0.28$, $p_{\mathrm{BIC}}(\mathrm{H1}|\mathrm{D}) = 0.57$ (see Fig. 18A). Bonferroni-corrected post hoc comparisons revealed that the massed group was more liberal than the control group ($p = .034$), but there were no other differences between the groups (p's > 0.426), and no other significant effects (p's > 0.259). Overall, the massed training group seemed to be more liberal in indicating the target was present (compared to the control group) up to 3 months following training.

2.2.3.1.4. Sensitivity. There were no significant effects for sensitivity comparing the post training session to performance 10 days after training for retention session (p's > 0.060) (see Fig. 17B). There were also no significant effects for sensitivity comparing post training to performance 3 months later for retention session (p's > 0.336) (see Fig. 18B).

2.2.3.1.5. Hits. There was a main effect of training group for hits 10 days later, $F(2,37) = 6.65$, $p = .003$, $\eta_p^2 = 0.26$, $p_{\mathrm{BIC}}(\mathrm{H1}|\mathrm{D}) = 0.91$. Bonferroni-corrected post hoc tests indicated that the adaptive and massed groups were better at finding the target 10 days after training compared to control group ($p = .047$, $p = .004$, respectively). Notably, this difference was still present, albeit not quite significant, 3 months following training, $F(2,20) = 3.35$, $p = .056$, $\eta_p^2 = 0.25$, $p_{\mathrm{BIC}}(\mathrm{H1}|\mathrm{D}) = 0.47$. LSD post hoc tests indicated that the adaptive and massed groups correctly detected the camouflaged target better than the control group 3 months after training (p's $= 0.035$). There were no other significant effects (p's > 0.154) (see Fig. 19).

2.2.3.1.6. False alarms. There were no significant effects for false alarms (p's > 0.258); overall, participants did not appear to differ in the frequency of false alarms for the forest images during retention (see Fig. 20).

2.2.3.1.7. Summary of forest background retention results. Overall, the forest background retention data results suggest that benefits for our training groups persisted 10 days and 3 months after training for both SDT and response times for the forest background images. Importantly, this benefit is restricted to situations where they correctly identified the target (i.e., hits), there were no benefits for reducing false alarms.

2.2.3.2. Urban backgrounds

2.2.3.2.1. Target present response times. There were no significant response time effects for the urban images on target present trials (p's $>$ 139; see Fig. 21).

2.2.3.2.2. Target absent response times. There were also no significant response time effects for the urban images on target absent trials (p's > 0.230; see Fig. 21).

2.2.3.2.3. Response criterion. There were also no significant response criterion effects for the urban images over retention (p's > 0.070; see Figs. 22 & 23).

2.2.3.2.4. Sensitivity. All participants were more sensitive to the urban targets 10-days later compared to immediately following training, $F(1,37) = 6.26$, $p = .017$, $\eta_p^2 = 0.15$, $p_{\mathrm{BIC}}(\mathrm{H1}|\mathrm{D}) = 0.80$. This finding is

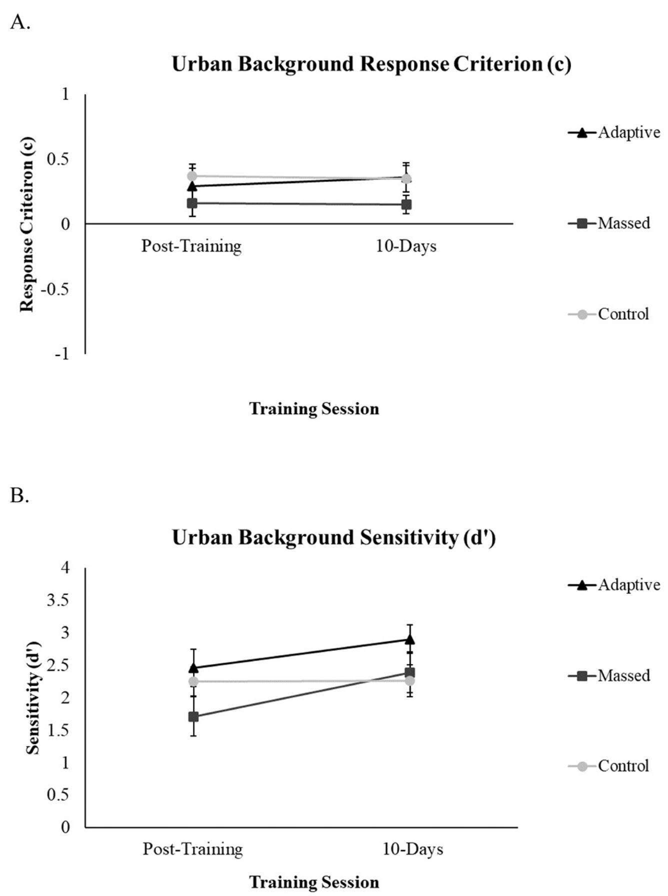

Fig. 22. Urban background 10-Day signal detection analyses by training group, response criterion (c) (A) and sensitivity (d') (B).
Note. Error bars represent the standard error of the mean. Adaptive N = 12, Massed N = 13, Control N = 15.

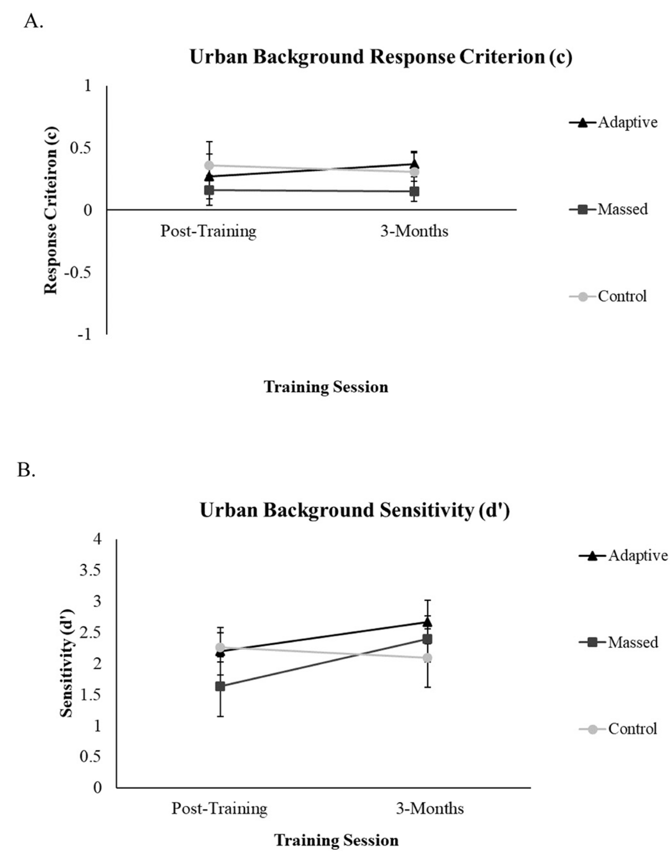

Fig. 23. Urban background 3-Months signal detection analyses by training group, response criterion (c) (A) and sensitivity (d') (B).
Note. Error bars represent the standard error of the mean. Adaptive $N = 8$, Massed N = 8, Control $N = 7$.

potentially indicative of a release for training for the forest images. However, this pattern did not persist at the 3-months session, $F(1,20) = 0.02$, $p = .880$, $\eta_p^2 < 0.01$, $p_{BIC}(H1|D) = 0.53$. There were no other significant effects (p's > 0.179).

2.2.3.2.5. Hits. All participants had more hits for the urban image targets 10 days later compared to immediately following training, $F(1,37) = 4.43$, $p = .042$, $\eta_p^2 = 0.11$, $p_{BIC}(H1|D) = 0.62$ (see Fig. 24). Speculatively, like the sensitivity results, this finding may suggest that all groups experienced a release from their training on the forest images. However, again this improvement did not persist 3 months later, $F(1,20) = 0.95$, $p = .340$, $\eta_p^2 = 0.05$, $p_{BIC}(H1|D) = 0.27$ (see Fig. 25). There were no other significant effects (p's > 0.394).

2.2.3.2.6. False alarms. There were no significant false alarm effects for the urban images over retention (p's > 0.301) (see Figs. 24 & 25).

2.2.3.2.7. Summary of urban retention results. Taken together, no interpretable results were found for the urban images 10 days and potentially 3 months after training. Although we found a possible release from transfer for sensitivity and hits at 10 days, this release was found for all three groups and did not persist at the 3-month session.

3. Discussion

We had four major goals in the present studies, (1) to characterize visual search training and transfer for camouflaged targets in more naturalistic settings, (2) to evaluate the extent to which any performance benefits are resilient to time-based atrophy after training, (3) to directly test the robustness of the perceptual learning hypothesis in explaining camouflage search improvements associated with training, and (4) to determine suggestions for future camouflage detection training paradigms that could be employed in real world contexts.

By utilizing naturalistic scenes for our search displays, we achieved our first goal; the search images used in our task are likely to contain highly similar scene statistics to those encountered in similar environments in the real world. Like previous research using less naturalistic scenes, transfer of training in search for camouflaged targets appears to extend to scenes that are novel (e.g., Chen & Hegdé, 2012; Neider et al., 2013), but similar to the scenes on which observers are trained.

To accomplish our second goal, we assessed the robustness of our two training groups' transfer performance over time. Transfer performance was assessed immediately following training, 10 days later, and 3 months later for both novel forest and urban scenes. Both of our camouflage training groups demonstrated faster response times and higher sensitivity relative to the control group immediately following training for novel forest images. Critically, this benefit persisted 10 days, and potentially even up to 3 months later, at least for hits. No meaningful differences were found between the groups for the urban images immediately following training, further supporting the finding that transfer of training in camouflage search appears to be somewhat specific. Overall, our retention data suggest that both our adaptive and massed camouflage search training methods engender performance benefits up to 3 months following training for similar natural images, but not to images portraying very different environments (i.e., urban scenes). Additionally, these training benefits appear to be specific to improving the ability to detect the presence of the target, rather than the absence of it.

Our third goal was to further examine the perceptual learning hypothesis within the context of naturalistic scenes. We initially predicted that if individuals are primarily learning underlying background scene

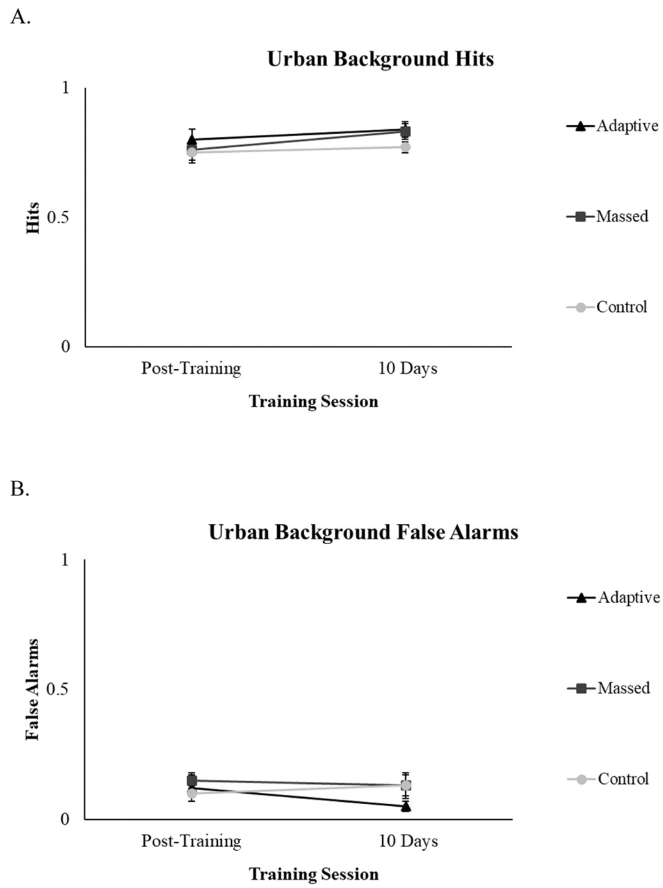

Fig. 24. Urban background 10-Day signal detection analyses by training group, hits (A) and false alarms (B).
Note. Error bars represent the standard error of the mean. Adaptive N = 12, Massed N = 13, Control N = 15.

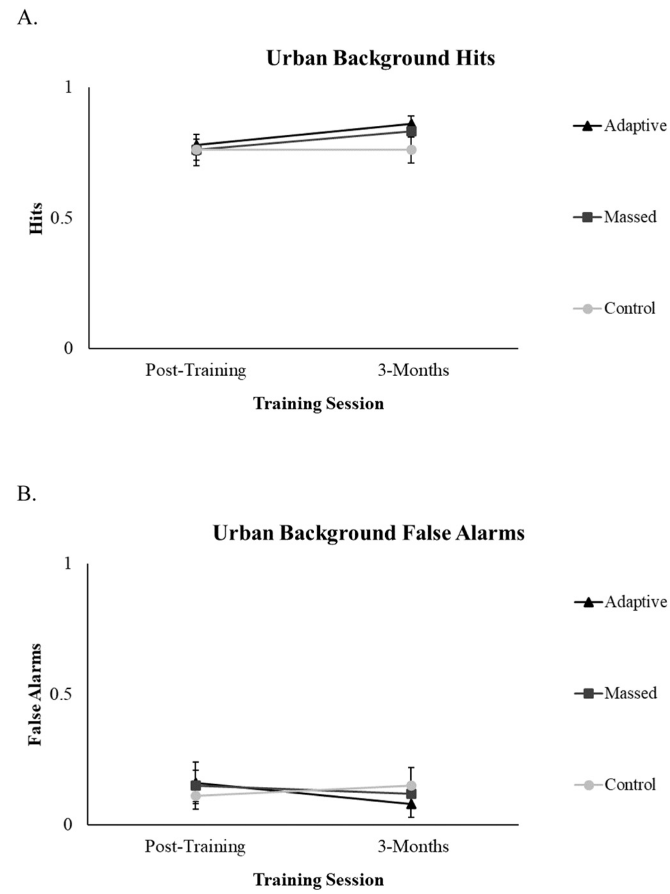

Fig. 25. Urban background 3-Months signal detection analyses by training group, hits (A) and false alarms (B).
Note. Error bars represent the standard error of the mean. Adaptive N = 8, Massed N = 8, Control N = 7.

statistics through training, and that is driving learning in search for camouflaged targets (as suggested by Chen & Hegdé, 2012), then they should exhibit search performance benefits for a range of targets provided the backgrounds they appear on are similar to those during training (i.e., forest scenes in our task). Somewhat surprisingly, we only found transfer of training for trials in which participants searched for target classes produced in the same manner as those used during training and in forest scenes. Neither of our training groups showed improved performance, relative to control, for our three novel target classes (i.e., motion blur, geometric, lens flare), or to urban scenes. Additionally, we observed training and transfer benefits for search through forest images, but no evidence that training on forest images transferred to urban images. Combined, these data provide some support for the perceptual learning hypothesis, but with a caveat. While it is reasonable, if not likely, that observers trained to search for camouflaged targets are indeed learning scene statistics associated with the backgrounds on which search scenes appear, our data suggest that this alone is unlikely to be the only mechanism underlying learning. In addition, it is likely that they are learning at least some information about the target and using that information to inform search processes as well. Importantly, this account is not incompatible with the perceptual learning hypothesis. It simply suggests that observers learn the scene statistics of *both* the background and target during training.

The final goal of the present study was to inform the development of future training paradigms for camouflaged detection in real world contexts. Our results center around three key suggestions for future camouflaged training: (1) which method of training results in the best performance, (2) how long this training lasts, and (3) how training transfers to novel scenes and targets. Regarding training methodology and efficacy, we found no differences for transfer of training between massed and adaptive camouflage search training groups. Although adaptive techniques are typically viewed as the gold standard, they are often compared against static difficulty levels that are often easier than the difficulty levels experienced in massed conditions (Von Bastian & Oberauer, 2014). Thus, it is possible that massed forms of training, in which individuals experience more challenging trials closer to transfer assessments, may offer a viable training method. Alternatively, it is also possible that if we had treated our adaptive training group differently, and had participants in that condition begin each training session at the highest difficulty level they achieved at their prior training session, adaptive training would have produced more robust performance improvements. Additional studies will be needed to further elucidate differences between these training approaches in search for camouflaged targets. However, for our current purposes, it is perhaps most critical to point out that training observers to detect camouflaged targets in natural images engendered transfer of training to novel, but similar, search displays. This finding suggests that a targeted training approach in which observers are trained to search for targets in similar contexts to their operational environments could prove effective in improving performance in those environments.

Our data also inform the extent to which training observers to search for camouflaged targets produces training benefits that are preserved over time. Specifically, we found that both of our camouflage training groups not only showed transfer of training effects relative to the control group, but that these performance benefits persisted long after training had ended. This is a particularly important finding, as training in the real world is not particularly useful if the performance benefits it produces are not preserved after training is completed.

The last suggestion for real world training paradigms concentrates on the transfer to novel targets and scenes. Our initial results suggest that any training benefits may be specific to not only the background trained upon, but also the targets presented. This means that any training methodology should incorporate operationally relevant scenes and target classes. For instance, militarily oriented training applications may be more focused on desert environments and lens flare targets; medically oriented training applications might focus on malignant tumors within mammograms. If multiple types of environments are likely to be encountered during operations (e.g., urban and forest environments, as used in the current study; or desert and vegetation-rich environments), then multiple training contexts may be necessary to engender performance improvements in all operational environments. It is worth noting that both of our training groups and the control group became better at the novel target classes (i.e., motion blur, geometric, lens flare) after training. This is potentially interesting given that our control group was only exposed briefly to the camouflage task before completing the novel target task. These findings suggest that our control participants benefited similarly to our training groups after only one recent camouflage search session. This is consistent with previous research (Hess et al., 2016) that indicates that robust performance improvements can be found after only one training session. What is difficult to predict is whether these performance improvements will persist after any meaningful duration of time. Although it is unlikely that widespread benefits should persist after such brief training, future work should explore the time-based atrophy of similar rapid training.

Declaration of competing interest

The authors report no conflict of interest.

Acknowledgements

This work was supported by Office of Naval Research Grant N000141512243.

References

Ahissar, M., & Hochstein, S. (1996). Learning pop-out detection: Specificities to stimulus characteristics. *Vision Research, 36*, 3487–3500.

Ball, K., & Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. *Science, 218*, 697–698.

Boot, W. R., Neider, M. B., & Kramer, A. F. (2009). Training and transfer of training in search for camouflaged targets. *Attention, Perception, and Psychophysics, 71*(4), 950–963.

Chen, X., & Hegdé, J. (2012). Learning to break camouflage by learning the background. *Psychological Science, 23*(11), 1395–1403.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*, 1087–1101.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Kit of Factor-Referenced Cognitive Tests*. Princteon, New Jersey: Educational Testing Services.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target in a nonsearch task. *Perception & Psychophysics, 16*, 143–149.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14*, 340–347.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality, 37*, 504–528.

Green, D. M., & Swets, J. A. (1988). *Signal detection theory and psychophysics*. Los Altos, CA: Peninsula Pub. (1988).

Hess, A. S., Wismer, A. J., Bohil, C. J., & Neider, M. B. (2016). On the hunt: Searching for poorly defined camouflaged targets. *PLoS One, 11*(3), Article e0152502.

Julesz, B. (1981). Textons, the elements of texture perception and their interactions. *Nature, 290*, 91–97.

Kristjánsson, A. (2015). Reconsidering visual search. *i-Perception, 6*(6), 1–14.

Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods, 43*, 679–690.

Motter, B., & Belky, E. (1998). The guidance of eye movements during active visual search. *Vision Research, 38*, 1805–1815.

Neider, M. B., Ang, C. W., Voss, M. W., Carbonari, R., & Kramer, A. F. (2013). Training and transfer of training in rapid visual search for camouflaged targets. *PLoS One, 8*(12), Article e83885.

Neider, M. B., Boot, W. R., & Kramer, A. F. (2010). Visual search for real world targets under conditions of high target–background similarity: Exploring training and transfer in younger and older adults. *Acta Psychologica, 134*(1), 29–39.

Neider, M. B., & Zelinsky, G. J. (2006a). Scene context guides eye movements during visual search. *Vision Research, 46*(5), 614–621.

Neider, M. B., & Zelinsky, G. J. (2006b). Searching for camouflaged targets: Effects of target-background similarity on visual search. *Vision Research, 46*(14), 2217–2235.

Palmer, E. M., Horowitz, T. S., Torralba, A., & Wolfe, J. M. (2011). What are the shapes of response time distributions in visual search? *Journal of Experimental Psychology: Human Perception and Performance, 37*(1), 58–71.

Raven, J., Raven, J. C., & Court, J. H. (2000). *Manual for Raven's standard progressive matrices*. San Antonio, TX: Pearson.

Reitan, R. M., & Wolfson, D. (1985). *The Halstead-reitan neuropsychological test battery*. Tucson, AZ: Neuropsychology Press.

Salthouse, T. A. (2010). Is flanker-based inhibition related to age? Identifying specific influences of individual differences on neurocognitive variables. *Brain and Cognition, 73*(1), 51–61.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27*, 763–776.

Sheldon, K. M. (1995). Creativity and self-determination in personality. *Creativity Research Journal, 8*, 61–72.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137–149.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97–136.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychological Review, 95*, 15–48.

Võ, M. L.-H. (2021). The meaning and structures of scenes. *Vision Research, 181*, 10–20.

Von Bastian, C. C., & Oberauer, K. (2014). Effects and mechanisms of working memory training: A review. *Psychological Research, 78*(6), 803–820.

Wechsler, D. (1997). *Wechsler adult intelligence scale* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Wolfe, J., Cave, K., & Franzel, S. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 419–433.

Wolfe, J. M. (1994a). Guided search 2.0: A revised model of guided search. *Psychonomic Bulletin and Review, 1*, 202–238.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science, 9*(1), 33–39.

Wolfe, J. M., Oliva, A., Horowitz, T. S., Butcher, S. J., & Bompas, A. (2002). Segmentation of objects from backgrounds in visual search tasks. *Vision Research, 42*, 2985–3004.

Zelinsky, G. J. (1996). Using eye saccades to assess the selectivity of search movements. *Vision Research, 36*, 2177–2187.