

Rapid Inversion of Singleton Distractor Representations Underlies Learned Attentional Suppression

Ziyao Zhang and Jarrod A. Lewis-Peacock

Department of Psychology, The University of Texas at Austin, Austin, Texas 78712

In visually complex and dynamically changing environments, humans must often filter out salient but task-irrelevant stimuli. Prior work shows that with repeated exposure to color singleton distractors, individuals can learn to divert attention away from these salient items. However, the neural mechanisms supporting such attentional suppression remain unclear. The present study examined the temporal trajectories of singleton distractor representations during visual search to address this gap. Using multivariate pattern analyses of EEG data in human subjects ($N = 40$, 30 females, 10 males), we identified two clusters of decodable singleton distractor representations: an early cluster from 100 to 200 ms and a later cluster from 200 to 400 ms. Temporal generalization analyses showed that the later representations were inverted versions of the early ones. Importantly, stronger late but not early representations predicted faster search responses, suggesting that the later signals support distractor suppression. This representational inversion facilitates suppressing singleton distractors in the spatial priority map. Comparing decoding evidence across locations revealed that singleton distractor locations were suppressed relative to nonsingleton distractors. Moreover, comparing the neural coding of locations revealed that the spatial organization in the singleton distractor neural space was inverted relative to that in the target neural space. Together, these findings reveal a rapid representational inversion underlying salient distractor suppression at the onset of visual search. This inversion of singleton distractor signals was likely driven by top-down control mechanisms that transform bottom-up saliency signals, producing an inverted arrangement of target and distractor information within a shared neural space.

Key words: attention suppression; distractor representations; MVPA; representational inversion; singleton distractor; visual search

Significance Statement

A primary goal of attention is to select relevant information while ignoring irrelevant input from the external environment. Although the mechanisms of attentional enhancement have been extensively studied, the mechanisms underlying attentional suppression remain less well understood. Here, we identify a representational transformation mechanism underlying learned suppression whereby neural representations of salient singleton distractors undergo a rapid inversion ~ 200 ms after search onset. These inverted representations can subsequently be read out as suppression signals during the computation of spatial priorities. Our findings suggest that transforming initial pop-out distractor signals into an inverted representational format might drive rapid attentional suppression and support goal-directed visual search.

Introduction

The ability to filter distracting information is crucial for goal-directed behavior. Salient but irrelevant stimuli, such as website pop-ups, readily capture attention, raising the question of whether such stimuli can be actively suppressed in attention. Recent progress on this issue has come from studies of singleton distractors

(e.g., a red item among green items) in visual search (Bacon and Egeth, 1994; Gaspelin et al., 2015). Color singletons capture attention and impair search performance (Jonides and Yantis, 1988; Theeuwes, 2010). However, accumulating evidence suggests that attentional capture by such distractors can be mitigated.

Received Oct. 8, 2025; revised Jan. 29, 2026; accepted Feb. 3, 2026.

Author contributions: Z.Z. designed research; Z.Z. performed research; Z.Z. contributed unpublished reagents/analytic tools; Z.Z. analyzed data; Z.Z. and J.A.L.-P. wrote the paper.

This work was completed with support from the National Institute of Health Grant R01EY028746 awarded to J.A.L.-P. This content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We acknowledge the invaluable contribution of Dr. Brad Stilwell, Dr. Howard

Egeth, and Dr. Nicholas Gaspelin for generously sharing their data. The dataset can be accessed at <https://osf.io/9xnsz/>.

The authors declare no competing financial interests.

Correspondence should be addressed to Ziyao Zhang at ziyaoz@utexas.edu.

<https://doi.org/10.1523/JNEUROSCI.1898-25.2026>

Copyright © 2026 the authors

Repeated exposure to color singleton distractors leads to faster detection of shape-defined targets when a singleton is present than when it is absent (Vatterott and Vecera, 2012; Gaspelin et al., 2015, 2017). Although this benefit has been interpreted as evidence for attentional suppression, alternative mechanisms have been proposed, including ignoring singletons in a serial search mode (Theeuwes, 2023), down-weighting the distractor feature dimension (Liesefeld and Müller, 2019, 2021), or rapid rejection of singleton distractors (Liesefeld et al., 2021, 2024).

Eye-tracking studies show that initial saccades are less likely to land on singleton distractors than on nonsingleton distractors (Gaspelin et al., 2017; Stilwell et al., 2023), suggesting suppression in overt attention. However, oculomotor measures cannot rule out rapid covert capture or fast reactive mechanisms (Klink et al., 2023). Converging electrophysiological evidence has identified the distractor positivity component (Pd) in conditions associated with behavioral suppression (Sawaki and Luck, 2010; Sawaki et al., 2012; Cosman et al., 2018). Although Pd has been linked to early inhibitory processes, its functional role remains debated (Gaspelin et al., 2023), with alternative accounts proposing that it reflects enhancement of nondistractor representations (Kerzel and Burra, 2020; Kerzel and Huynh Cong, 2023) or sensory imbalances between visual hemifields (Oxner et al., 2025).

These findings have motivated the signal suppression hypothesis, which posits that salient stimuli automatically generate bottom-up priority signals that can be suppressed by top-down mechanisms before attention is deployed (Sawaki and Luck, 2010; Gaspelin and Luck, 2018b). In this framework, suppression operates by down-weighting the distractor feature prior to any attentional shift (Gaspelin and Luck, 2018a; Gaspelin et al., 2025). In contrast, the rapid disengagement hypothesis proposes that suppression occurs reactively, after attention has been captured by the distractor (Theeuwes, 2010). Recent computational modeling suggests that these accounts may reflect differences in the relative timing of target- and distractor-related priority signals rather than distinct mechanisms (Zhang et al., 2025).

Despite advances in behavioral and computational accounts, the neural mechanisms underlying attentional suppression remain unclear. One possibility is neural suppression, whereby distractor-related activity or representations are reduced during visual search (Turatto et al., 2018; Won et al., 2019). Consistent with this view, fMRI and nonhuman primate studies have reported reduced neural responses to repeated singleton distractors in the early visual cortex and in priority map regions, including the intraparietal cortex, frontal eye fields (FEF), and lateral intraparietal cortex (LIP; Ipata et al., 2006; Cosman et al., 2018; Won et al., 2020; Adam and Serences, 2021).

In contrast, recent evidence suggests that attentional suppression may rely on mechanisms other than uniform neural suppression. Neuronal populations in V4, FEF, and LIP exhibit heterogeneous response profiles to singleton distractors rather than consistent reductions in activity (Klink et al., 2023; Sapountzis et al., 2025). Moreover, decoding analyses indicate that distractor representations remain stable with training, raising the possibility that these representations are transformed into a format that no longer interferes with attentional allocation to targets (Sapountzis et al., 2025).

Here, we tested this representational transformation account using temporal generalization analyses of EEG data. Whereas neural suppression accounts predict that distractor representations weaken over time, transformation accounts predict changes in representational format without necessarily reduced strength. Consistent with the latter, we observed reliable representations of

singleton distractor locations across time. Notably, the representational format transformed ~ 200 ms after search onset, with later representations reflecting an inverted version of the early representations.

Materials and Methods

We reanalyzed data shared from Stilwell et al. (2022). Two groups of 20 participants were recruited for Experiment 1 and Experiment 2, respectively ($N = 40$, 30 females, 10 males). In each trial, participants searched for a predefined target item (circle or diamond, counterbalanced between participants) in a search array. In the search array, the inner ring contained a target shape and distractors that were in different shapes. On 75% of trials, a color singleton distractor that had a distinct color from other items was presented in the inner ring. The color singleton distractor was never the target, and participants were instructed to use the shape feature to locate the target item. The specific color of the color singleton distractor (red or green) was fixed for each participant but counterbalanced across participants. The outer ring contained nontarget shapes only to boost the relative salience of the singleton distractor. Overall, participants completed 1,296 singleton-present trials and 432 singleton-absent trials. Search arrays in Experiment 1 had an inner ring with four items (effective set size 4), while in Experiment 2, the inner ring had eight items (effective set size 8). EEG signals were collected and preprocessed following the procedures outlined in Stilwell et al. (2022).

Decoding of target and singleton distractor locations. Location decoding was conducted separately for singleton and singleton-absent trials. For singleton-absent trials, we used support vector machine (SVM) to classify the target item's location based on the spatial distribution of the EEG signal across 17 posterior electrodes (Pz, P3, P5, P7, P9, PO7, PO3, O1, POz, Oz, P4, P6, P8, P10, PO4, PO8, O2). We implemented this model using the *svc* function from the *sklearn* package in Python. EEG data were downsampled to 100 Hz, and decoding was performed for each time point (10 ms) ranging from -200 to 800 ms relative to the onset of the search array. For each time point, the dataset was divided into threefold. An equal number of trials from each location class were randomly selected for each fold without replacement to ensure unbiased training. Two folds were used as the training set, and the remaining fold served as the testing set. This procedure was repeated for the threefold, with each one serving as the testing set in an iteration. The entire process was repeated 100 times (Wolff et al., 2020), and decoding accuracies were averaged across 100 iterations. Decoding of target and singleton distractor locations in singleton-present trials followed the same procedure as decoding target locations in singleton-absent trials, except for adjusting the training set to match the number of trials in singleton-absent trials. This adjustment aimed to ensure equal training data for decoders in both singleton and singleton-absent trials. This should prevent biases in decoding performance comparisons between singleton and singleton-absent trials resulting from uneven training sets.

Linear models linking decoding evidence to reaction time. To link target location representations to reaction time, decoding evidence for the correct target location in each trial was obtained from the *decision_function* in *sklearn*, which reflects the model's confidence (distance to the decision hyperplane). Higher values indicate stronger neural representations of the target location. Decoding evidence was averaged within the early (100–200 ms) and later (200–400 ms) time windows. Both the averaged decoding values and reaction times were *z*-scored within participants before fitting the linear models. Linear mixed-effect models were then fitted using the *lmer* function in R, with early and late decoding evidence as predictors and reaction time as the outcome variable. To link singleton distractor representations to reaction time, we applied the same analysis pipeline, except that decoding evidence was extracted for singleton distractor locations rather than target locations.

Latency estimation. To estimate the onset latency of above-chance decoding, we applied a jackknife-based bootstrapping procedure

(Miller et al., 1998; Ulrich and Miller, 2001). We generated N bootstrapped subsamples ($N = 20$ in both experiments), each created by leaving out one participant. For each subsample, we performed a one-sample t test comparing decoding accuracy to chance. The latency for that subsample was defined as the onset of the earliest cluster of significant time points. This procedure yielded N latency estimates. These estimates were then used for statistical comparisons.

Temporal generalization analysis. Temporal generalization analyses were conducted for target locations in singleton-absent trials and singleton locations in singleton-present trials. The same decoding procedures were used, with the only difference being that trained decoders for each time point were used to predict location labels for data at all time points.

Cross-condition generalization. Cross-condition generalization was performed between target locations in singleton-absent trials and singleton distractor locations in singleton-present trials. SVM decoders were trained exclusively on data from singleton-absent trials with labels indicating target locations. These trained decoders were then used to predict singleton distractor locations in singleton-present trials. Decoding accuracies were computed based on the probability of the decoders' predictions matching the actual singleton distractor location labels.

Activation scores. To compute activation scores for each location in singleton-present trials, we first trained decoders on target locations in singleton-absent trials. These trained decoders were subsequently used to predict location labels in singleton-present trials. Instead of generating prediction labels, prediction probabilities were produced to indicate the likelihood of a given location being the true label based on the neural patterns. Prediction probabilities for target locations, singleton locations, and nonsingleton distractor locations were averaged across trials for each subject. For plotting the spatial priority map, the averaged prediction probabilities of nonsingleton distractor locations (two in Experiment 1, six in Experiment 2) were used as the baseline activation. The activation for target and singleton distractor locations was computed as the difference between the prediction probabilities of these locations and the baseline.

Correlations between target and singleton representations. Raw EEG data of singleton-absent trials (200–400 ms) with the same target location were averaged to create the mean target representations for each potential target location (four in Experiment 1 and eight in Experiment 2). Likewise, singleton-present trials with the same singleton distractor location were averaged (200–400 ms) to generate singleton distractor representations for all potential singleton distractor locations (four in Experiment 1 and eight in Experiment 2). Pearson's correlations in multichannel EEG activities between target and singleton representations were computed for each potential location and then averaged. Positive correlations indicate similarity between the location representations in the target space and location representations in the singleton distractor space, while negative correlations suggest inverted representations.

Principal component analysis (PCA) was utilized to visualize the relationship between the target representational space and the singleton distractor representational space. For this purpose, we identified the first two principal components (PCs) of the target representational space for one example participant (99% variance explained). Both target and singleton distractor representations were projected into the space defined by these two PCs. It should be noted that PCA was used to visualize the relationship between the target and singleton distractor representation only. The formal correlation analysis was performed based on the raw signal space.

Statistics. We conducted one-sample (one-sided) t tests to compare decoding accuracies against the theoretical chance level, as below-chance decoding is not meaningful for this analysis. All other comparisons were performed using two-sided t tests. One-sample t tests were also used to compare temporal generalizability scores and cross-condition generalizability scores against the theoretical chance level. In both cases, above-chance scores indicated positive generalizability, whereas below-chance scores indicated negative generalizability. Paired t tests were used to

compare prediction probabilities for target locations, singleton distractor locations, and nonsingleton distractor locations. Finally, one-sample t tests were conducted to compare neural correlations between target and singleton distractor representations against zero, with positive values indicating positive correlations and negative values indicating negative correlations. All analyses were performed in Python using standard statistical packages. Bayes factors for the alternative hypothesis (BF10) were reported, with values between 1 and 3 indicating anecdotal evidence, values between 3 and 10 indicating moderate evidence, and values greater than 10 indicating strong evidence. To account for multiple comparisons, Sidak correction was applied, and adjusted p values were reported.

Results

Participants searched for a predefined target item (circle or diamond, counterbalanced between participants) in a search array. In the search array, the inner ring contained a target shape and distractors that were in different shapes. On 75% of trials, a color singleton distractor that had a distinct color from other items was presented in the inner ring. The color singleton distractor was never the target, and participants were instructed to use the shape feature to locate the target item. The outer ring contained nontarget shapes only to boost the relative salience of the singleton distractor.

Singleton distractor representations showed distinct temporal patterns compared with target representations

We applied multivariate pattern analysis (MVPA) to track target and singleton distractor representations across time. Both singleton distractor and target locations showed successful decoding (Fig. 1). Target location representations emerged ~ 200 ms and persisted for most of the search period. Interestingly, the decoding of singleton distractor locations revealed an earlier initial peak, spanning from 100 to 200 ms. This initial peak was followed by a subsequent, more robust peak occurring between 200 and 400 ms for both set size 4 and set size 8. The decoding accuracies of singleton distractor locations quickly dropped off following the second peak whereas that for target locations persisted. The estimated latency for singleton distractor representations was faster than that for target representations in both singleton-present trials (Experiment 1, $t_{(19)} = -5.17$; $p < 0.001$; $d = 1.58$; BF10 = 477.01; Experiment 2, $t_{(19)} = -259.0$; $p < 0.001$; $d = 81.90$; BF10 = 1.2×10^{31}) and singleton-absent trials (Experiment 1, $t_{(19)} = -9.92$; $p < 0.001$; $d = 3.15$; BF10 = 2.0×10^6 ; Experiment 2, $t_{(19)} = -46.15$; $p < 0.001$; $d = 14.59$; BF10 = 4.3×10^{17}). Interestingly, the latency for target representations in singleton-present trials was faster than in singleton-absent trials (Experiment 1, $t_{(19)} = -11.86$; $p < 0.001$; $d = 3.87$; BF10 = 3.1×10^7 ; Experiment 2, $t_{(19)} = -12.11$; $p < 0.001$; $d = 3.79$; BF10 = 4.4×10^7).

The identification of two early peaks in singleton distractor decodings suggests potential shifts in representation during the search process. We conducted temporal generalization analyses to further examine these potential changes. Our methodology involved training decoders at one time point and then testing the decoder across all time points, repeating this procedure for each time point. As shown in Figure 2, target representations were stable over time. Trained decoders were able to generalize to neighboring time points. In contrast, the temporal generalization analysis of singleton distractor representations showed two discernible clusters. The initial cluster extended approximately from 100 to 200 ms, while the subsequent cluster emerged at ~ 200 ms. Crucially, these two clusters showed an inverted relationship. Decoders trained on activity from 100 to 200 ms had below-chance decoding performance when tested on data from

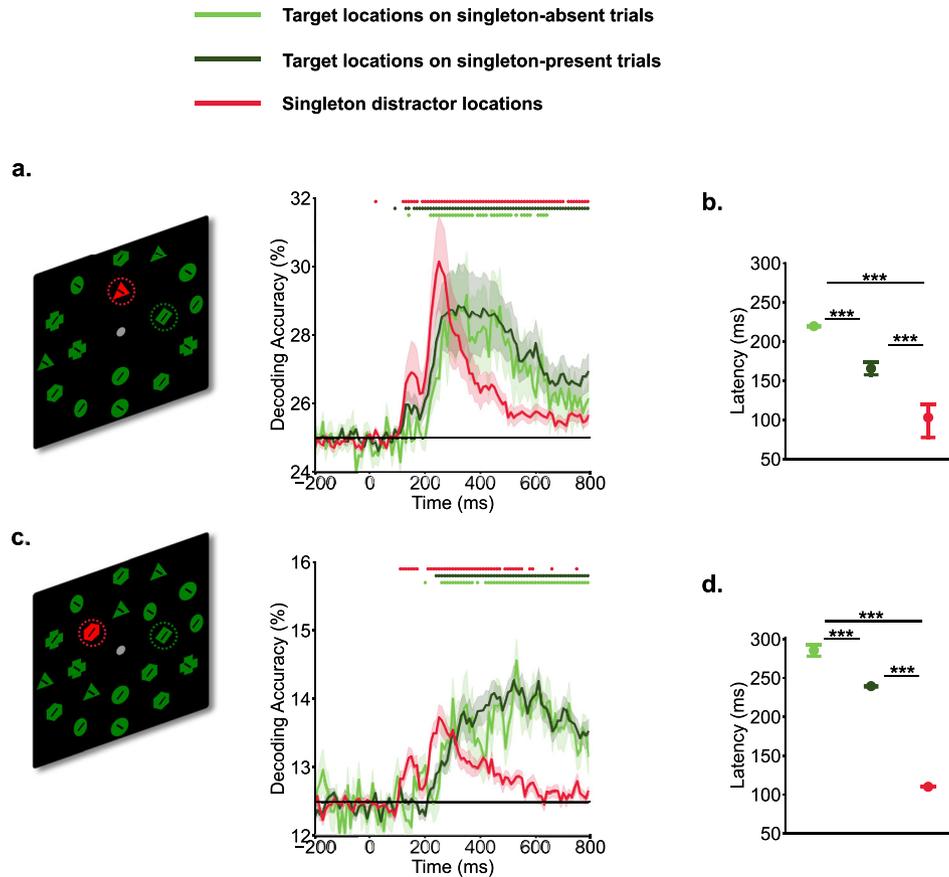


Figure 1. Temporal profiles of target and singleton distractor representations. Participants searched for a specific target shape (e.g., a green diamond) within the inner ring of a search array and made a speeded button press to indicate the direction of the line within the target shape (left vs right). Time 0 ms represents the onset of search arrays. In Experiment 1 (*a* set size 4) and Experiment 2 (*b* set size 8). *a, c*, Decoding of singleton distractors showed an earlier initial peak of decoding evidence from 100 to 200 ms, followed by a later peak from 200 to 400 ms. Target decoding showed a gradual increase in evidence. The horizontal black line indicates chance-level (25% for set size 4 and 12.5% for set size 8). Shaded areas indicate standard errors. Colored dots indicate above chance-level decoding performance ($p < 0.05$, one-sided t test). *b, d*, Latency of decoding accuracies estimated via the jackknife-based procedure. *** indicates $p < 0.001$.

200 to 400 ms and vice versa. This negative generalizability performance suggests an inversion of representations for singleton distractors between the early (100–200 ms) and the later (200–400 ms) time windows. It also provides preliminary evidence for the signal transformation hypothesis that neural codings of singleton distractors were inverted during the search. However, it is unclear yet how this inversion directly supported attentional suppression.

We linked trial-by-trial decoding evidence for target and singleton distractor locations to reaction time to examine their functional roles. The relationship between early (100–200 ms) and late (200–400 ms) singleton distractor representations and reaction time is particularly informative for distinguishing between theoretical accounts of distractor processing. Under the reactive suppression (rapid disengagement) framework, attentional suppression occurs after attention has been captured by the singleton. Accordingly, early distractor representations would be expected to reflect attentional capture, whereas later representations would reflect reactive suppression. Based on this account, early distractor representations should be associated with slower reaction times, whereas later representations should be associated with faster reaction times. In contrast, under the proactive suppression framework (signal suppression), both early and late distractor representations likely reflect different stages of active suppression. Under this framework, both representations,

particularly the later ones, should be associated with faster reaction times. Our linear modeling results (Fig. 3) showed that only late target representations negatively predicted reaction times in both singleton-absent (Experiment 1, $\beta = -0.04$; CI = $[-0.06, -0.02]$; $p = 0.001$; BF10 = 11.22; Experiment 2, $\beta = -0.06$; CI = $[-0.08, -0.03]$; $p < 0.001$; BF10 = 76.21; combined, $\beta = -0.05$; CI = $[-0.06, -0.03]$; $p < 0.001$; BF10 = 21,555) and singleton-present trials (Experiment 1, $\beta = -0.07$; CI = $[-0.08, -0.05]$; $p < 0.001$; BF10 = 2.3×10^{20} ; Experiment 2, $\beta = -0.08$; CI = $[-0.10, -0.06]$; $p < 0.001$; BF10 = 2.1×10^{29} ; combined, $\beta = -0.07$; CI = $[-0.08, -0.06]$; $p < 0.001$; BF10 = 6.7×10^{50}), whereas early target representations did not ($ps > 0.05$). Critically, early singleton distractor representations did not predict reaction times (Experiment 1, $\beta = 0.00$; CI = $[-0.01, 0.01]$; $p = 0.942$; BF10 = 0.009; experiment 2: $\beta = 0.01$; CI = $[-0.00, 0.03]$; $p = 0.083$; BF10 = 0.021; combined, $\beta = 0.01$; CI = $[-0.00, 0.02]$; $p = 0.199$; BF10 = 0.011), whereas late distractor representations negatively predicted reaction times (Experiment 1, $\beta = -0.02$; CI = $[-0.03, 0.00]$; $p = 0.008$; BF10 = 65.79; Experiment 2, $\beta = -0.02$; CI = $[-0.03, -0.00]$; $p = 0.035$; BF10 = 3.52; combined, $\beta = -0.02$; CI = $[-0.03, -0.01]$; $p = 0.001$; BF10 = 11.45). Together, these findings indicate that both target and singleton distractor representations in the late window are associated with faster reaction times and thus enhanced behavioral performance. Distractor representations in the early time window did not predict

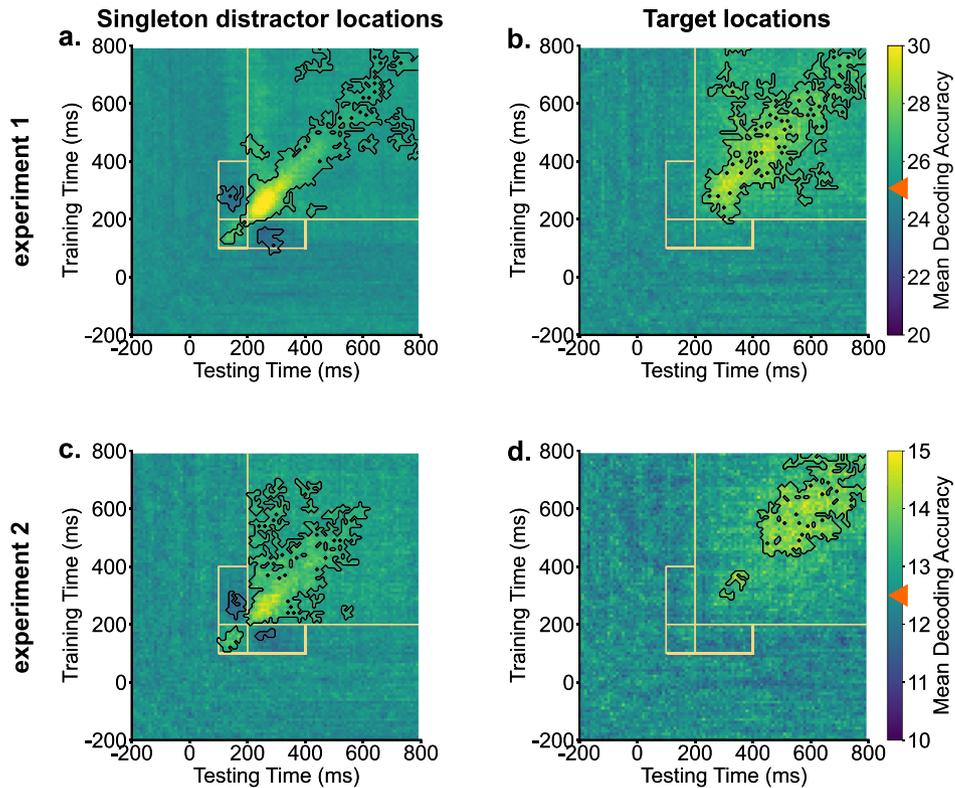


Figure 2. Singleton distractor representations showed inverted transformations, whereas target representations remained stable. Across experiments, generalization analyses of singleton distractor representations (*a, c*) revealed two clusters. Training decoders on activities from 100 to 200 ms resulted in below chance-level decoding evidence when tested on activities from 200 to 400 ms and vice versa. Target representations demonstrated stability and could be generalized to neighboring time points (*b, d*). Orange boxes indicate key generalization time windows. Black contours indicate significant areas ($p < 0.05$). Bright colors indicate above chance-level decoding performance (denoted by orange triangles), and dark colors indicate below chance-level decoding performance.

behavioral performance, which runs counter to a rapid disengagement interpretation of our results. The signal suppression hypothesis could potentially account for this pattern but only if the early component of distractor processing reflects an initial registration of salience that is subsequently used for diverting attention away. However, this interpretation would be less consistent with a strict proactive suppression framework, which posits that suppression is implemented prior to the onset of the search array.

Singleton distractor locations were suppressed in the spatial priority map

To investigate how the inversion of singleton representations supported attentional suppression and facilitated visual search, we performed cross-condition generalization analyses. In this procedure, decoders were trained on target representations in singleton-absent trials, where top-down enhancements served as the primary guiding sources to boost target locations in the priority map. Subsequently, these trained decoders were tested on singleton-present trials to reveal the priority maps in instances where both target enhancement and singleton suppression might influence spatial priorities. The cross-condition generalization analyses revealed that singleton distractor locations were suppressed in the spatial priority map, resulting in below chance-level decoding performance (Fig. 4*a,b*). This suppression of singleton distractor locations started ~ 200 ms and persisted. Intriguingly, no generalization was observed between target

representations and singleton distractor locations during the early time window from 100 to 200 ms, corresponding to the initial cluster of singleton distractor representation observed in temporal generalization analyses.

Below-chance decoding performance could reflect suppression of singleton distractor locations. Alternatively, it might be driven by a strong enhancement of target locations, which would reduce decoding performance for all other locations. To rule out the alternative possibility, we compared the spatial priority of singleton distractors against a baseline of nonsingleton distractors in the search array. We derived prediction probabilities from trained decoders for each location in the search array and compared them across target locations, singleton distractor locations, and nonsingleton distractor locations. Activation scores for target and singleton distractor locations were computed as the difference in prediction probabilities compared with the baseline (nonsingleton distractor locations; for one example time point, see Fig. 4*c*). For between 200 and 400 ms into the search, we found enhancement of target locations compared with the baseline (Experiment 1, $t_{(19)} = 2.14$; $p = 0.045$; $d = 0.81$; $BF_{10} = 1.51$; Experiment 2, $t_{(19)} = 2.93$; $p = 0.009$; $d = 1.02$; $BF_{10} = 5.78$). Critically, suppression of singleton distractor locations compared with the baseline was also observed across the experiment (Experiment 1, $t_{(19)} = -2.11$; $p = 0.048$; $d = 0.32$; $BF_{10} = 1.42$; Experiment 2, $t_{(19)} = -2.62$; $p = 0.017$; $d = 0.68$; $BF_{10} = 3.30$). As shown in Figure 4, *d* and *e*, prediction probabilities of target locations displayed enhancement relative to the baseline from 200 ms

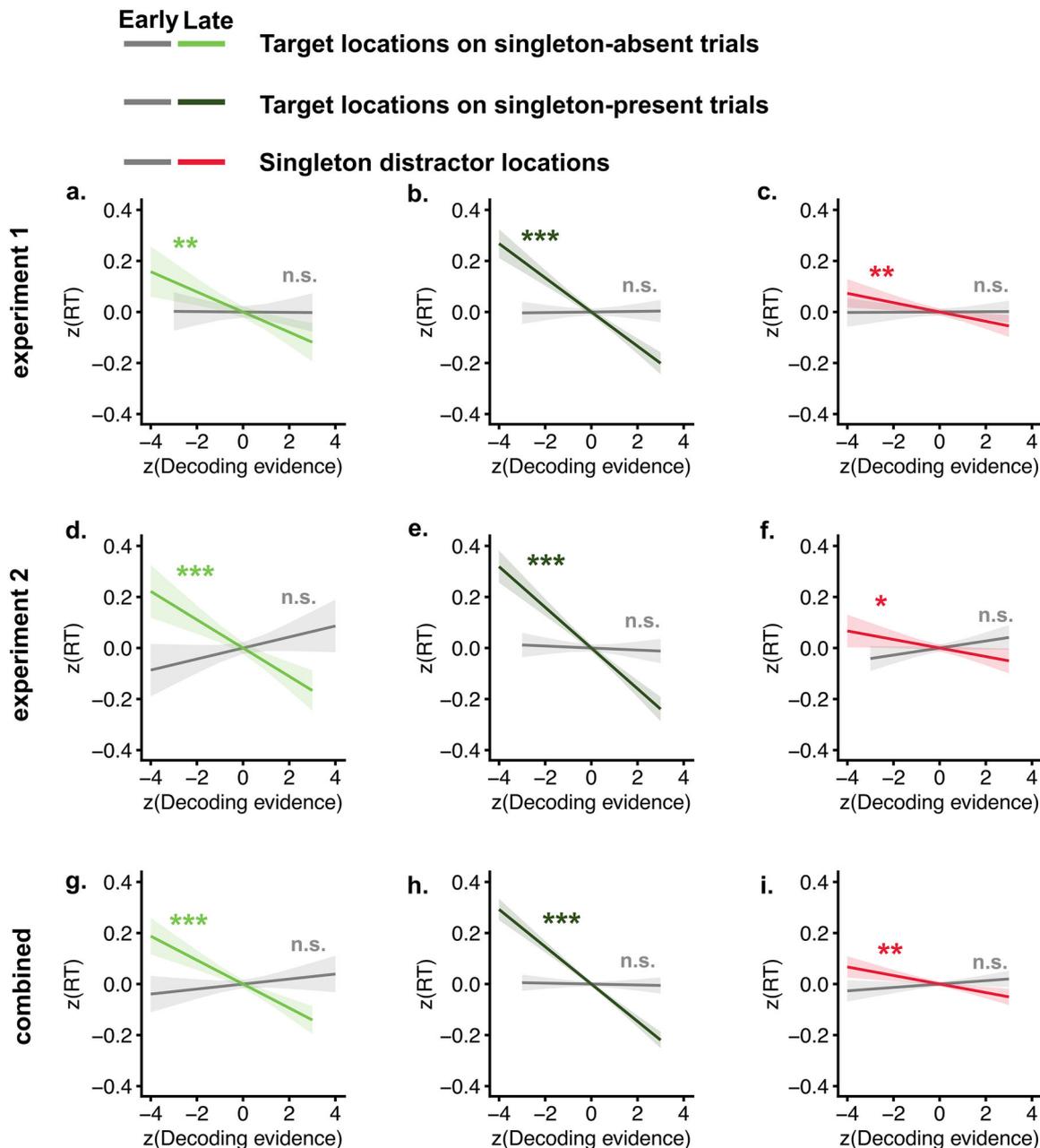


Figure 3. Target and singleton distractor representations predict faster reaction times. **a–c**, Target representations in singleton-absent, singleton-present, and singleton distractor representations predicted reaction time in Experiment 1. Gray indicates the early window (100–200 ms); light green, dark green, and red indicate the late window (200–400 ms). **d–f**, The linear model results for Experiment 2. **g–i**, Linear model results combining Experiments 1 and 2. Shaded areas represent confidence intervals. n.s. indicates nonsignificant. * indicates $p < 0.05$; ** indicates $p < 0.01$; *** indicates $p < 0.001$.

onward. In contrast, prediction probabilities of singleton distractor locations were suppressed below the baseline.

Singleton distractor and target locations were coded in a shared neural space but in a reversed manner

Cross-condition generalization analyses suggested that singleton distractors were suppressed in the neural space representing target locations. We hypothesized that this suppression might be driven by coding distractor information in an inverted format relative to target information. Such inverted coding of target and distractor information could facilitate the integration of top-down enhancement and suppression in computing spatial priority (Itti and Koch, 2000; Fecteau and Munoz, 2006; Aw

et al., 2012; Luck et al., 2021). To directly test this hypothesis, we conducted neural correlation analyses to examine the relationship between target and singleton distractor representations.

In Figure 5a, representations of both targets and singletons were shown for an example subject from 200 to 400 ms into the search, projected in a 2D subspace for easier visualization. This subspace was constructed using the first two PCs (a total of 99% variance explained) of the neural representations of target locations. The location information demonstrated clear organization; the responses to four locations were well separated, and the distance between neural representations corresponded closely to the physical distance between locations (i.e., neighboring locations had neighboring representations). Interestingly,

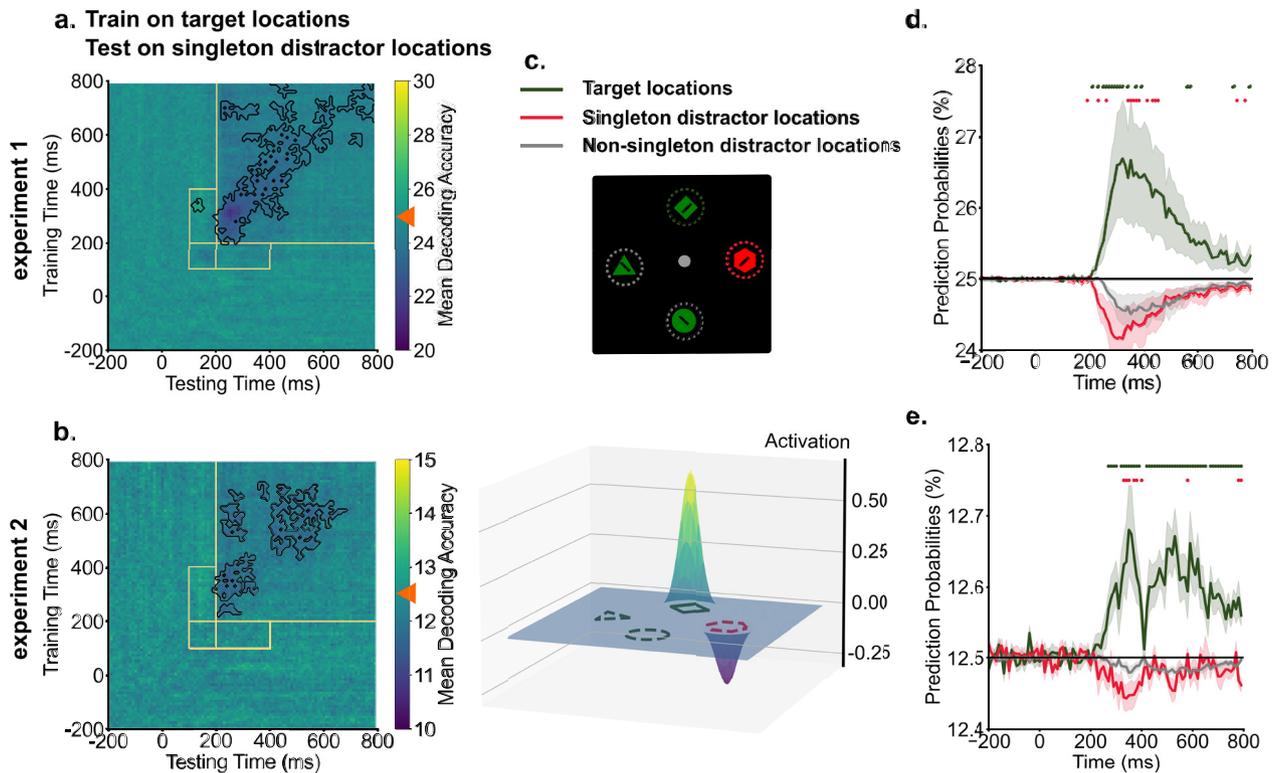


Figure 4. Singleton distractor locations were suppressed in the spatial priority map. **a, b.** Across experiments, training decoders on target locations from singleton-absent trials demonstrated below chance-level decoding performance (denoted by orange triangles) for singleton distractor locations, indicating the suppression of singleton distractor locations. Dark colors indicate below chance-level decoding performance for singleton locations. Black contours indicate significant areas ($p < 0.05$). **c.** An example (250 ms in Experiment 1) illustrates how we computed activation scores from decoding prediction probabilities. Prediction probabilities of target locations (green) and singleton distractor locations (red) were baselined to the averaged prediction probabilities of nonsingleton distractor locations (gray). **d, e.** The time course of prediction probabilities for target locations and singleton distractor locations were plotted against the baseline (other locations). From 200 ms into the search, target locations showed increased prediction probabilities, whereas singleton distractor locations showed decreased prediction probabilities compared with the baseline. Shaded areas indicate standard errors. Colored dots indicate significant differences in prediction probabilities between target and singleton distractor locations to nonsingleton distractor locations ($p < 0.05$).

singleton distractor and target representations showed a reversed pattern. For instance, the top location (blue square) in the target representational space was projected to the top-right corner, whereas the same location in the singleton distractor representational space was projected to the bottom-left region (blue circle).

Formal correlation analyses were performed to link target representations and singleton distractor representations in the raw signal space (dimension, channel). For between 200 and 400 ms into the search, target and singleton distractor representations showed reliable negative correlations (Experiment 1, $t_{(19)} = -3.22$; $p = 0.005$; $d = 0.72$; $BF_{10} = 9.99$; Experiment 2, $t_{(19)} = -3.05$; $p = 0.006$; $d = 0.68$; $BF_{10} = 7.29$). As shown in Figure 5, *b* and *c*, negative correlation emerged ~ 200 ms and persisted into the later search period, suggesting that singleton distractor representations and target representations were coded with an inverted representational geometry.

Discussion

Recent attention research has increasingly shifted from an exclusive focus on target enhancement to a broader consideration of distractor suppression (Geng et al., 2019; Luck et al., 2021; Wöstmann et al., 2022; Theeuwes, 2025). Although converging evidence indicates that attention can be directed away from cued or learned distractors (Sawaki and Luck, 2010; Arita et al., 2012; Vatterott and Vecera, 2012; Gaspelin et al., 2015, 2017; Chang and Egeth, 2019; Won et al., 2019; Zhang et al., 2020,

2022; Zhang and Carlisle, 2023), it remains unclear how suppression unfolds over time and how to-be-suppressed information is represented neurally. The primary goal of the present study was to test a signal transformation mechanism underlying singleton distractor suppression during visual search. Using MVPA of EEG data, we characterized the temporal evolution of singleton distractor representations.

We found that singleton distractor representations emerged reliably earlier than target representations, consistent with the idea that salient distractors generate rapid bottom-up signals prior to the emergence of top-down target enhancement (Jonides and Yantis, 1988; Theeuwes, 2010). Increasing the set size from four to eight items delayed the emergence of target representations but did not affect singleton distractor representations, in line with prior work showing that crowding weakens target signals (Reddy and VanRullen, 2007), while salience-driven signals remain relatively stable. Interestingly, target representations emerged earlier in singleton-present than singleton-absent trials. Although this pattern could suggest facilitation of target processing by the presence of a singleton distractor, it may also reflect a limitation of spatial decoding in displays where target and singleton locations are mutually exclusive, as discussed below.

Decoding analyses revealed two distinct peaks in singleton distractor representations. Temporal generalization analyses showed that target representations were stable across time, forming a sustained cluster from ~ 200 ms onward. In contrast,

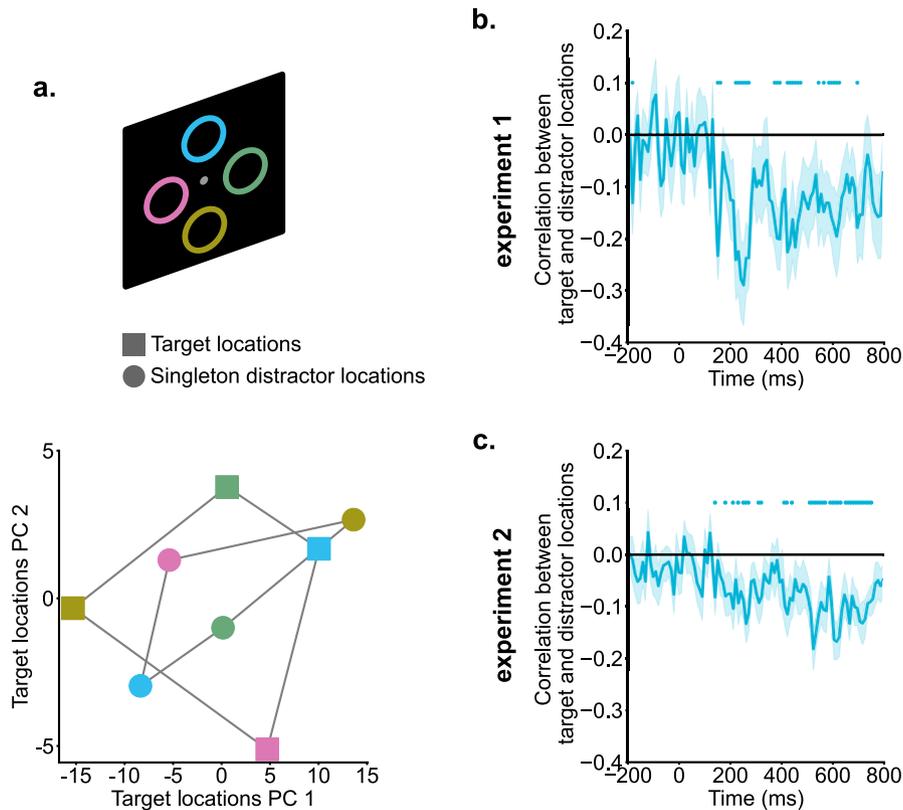


Figure 5. Singleton distractor suppression and target enhancement relied on inverted neural coding. *a*, Singleton distractor and target representations of an example subject were projected onto a 2D target representational space for illustration purposes. Different colors indicate different locations. Singleton distractor and target representations exhibited an inverted pattern. *b*, *c*, Correlations between singleton distractor and target representations. Starting from ~ 200 ms, singleton distractor and target representations showed negative correlations. Shaded areas indicate standard errors. Colored dots indicate significant correlations compared with zero ($p < 0.05$).

singleton distractor representations formed two temporally dissociable clusters: an early cluster from 100 to 200 ms and a later cluster ~ 200 –400 ms. Critically, these two singleton clusters exhibited negative cross-temporal generalization, indicating a representational inversion ~ 200 ms. This pattern provides direct evidence that singleton distractor representations undergo a transformation rather than a monotonic attenuation.

To assess how this representational inversion influenced spatial priority computations, we trained decoders on target locations from singleton-absent trials, where spatial priority is primarily driven by top-down target enhancement, and applied them to singleton-present trials. This cross-condition generalization yielded below-chance decoding at singleton distractor locations. Such below-chance decoding could reflect either enhanced target representations, which would push prediction probabilities for all nontarget locations below chance, or suppression specifically targeting the singleton location. To distinguish between these possibilities, we compared prediction probabilities at singleton distractor locations with those at nonsingleton distractor locations. Target locations showed above-baseline enhancement beginning ~ 200 ms, whereas singleton distractor locations showed reliable below-baseline suppression over the same time window. These findings provide converging evidence that singleton distractor representations are actively suppressed within the spatial priority map.

A key limitation of spatial decoding is that target and singleton distractor locations are inherently anticorrelated. Consequently,

negative decoding evidence could arise from either suppression of a specific location or enhancement elsewhere. Although our comparison against nonsingleton distractor locations partially mitigates this concern, it cannot fully resolve this ambiguity. Future work should test suppression using complementary approaches, such as feature-based decoding or frequency tagging (Forschack et al., 2017; van Moorselaar et al., 2020), particularly in paradigms where target and distractor features vary independently.

We hypothesized that suppression is implemented by coding singleton distractors in an inverted representational format relative to targets. Such inverted coding would facilitate downstream integration of target enhancement and distractor suppression within a shared neural substrate. Consistent with this hypothesis, correlations between target and singleton distractor representations revealed robust negative relationships beginning ~ 200 ms, indicating that target and singleton distractor locations were inversely coded within a shared neural space. This shared space likely underlies the computation of spatial priorities, allowing target and distractor signals to exert opposing influences. While these findings support the existence of a common coding framework, further work is needed to more precisely characterize the relationship between target and distractor subspaces and their link to the hypothesized priority map.

A substantial body of behavioral, eye-tracking, electrophysiological, EEG, and fMRI evidence supports successful suppression of salient color singleton distractors (Sawaki and Luck, 2010; Sawaki et al., 2012; Vatterott and Vecera, 2012; Gaspelin et al.,

2015, 2017; Cosman et al., 2018; Chang and Egeth, 2019; Stilwell et al., 2022). Electrophysiological recordings in NHPs and fMRI evidence in humans have shown reduced neuronal responses to singleton distractors relative to nonsingleton distractors (Ipata et al., 2006; Cosman et al., 2018; Won et al., 2020; Adam and Serences, 2021). These findings have been interpreted as evidence for signal attenuation.

However, recent work challenges a purely attenuation-based account. Neural responses to singleton distractors appear highly heterogeneous in V4, FEF, and LIP (Klink et al., 2023; Sapountzis et al., 2025). Moreover, decoding analyses indicate that singleton distractor representations persist rather than vanish, suggesting that they may play an active role in guiding behavior. Together, these findings suggest that singleton distractor suppression may rely on representational transformation rather than uniform attenuation.

The present study extends this work by providing direct evidence for a rapid inversion of singleton distractor representations during visual search. We propose that attentional suppression begins with early sensory registration of salient singleton distractors, reflected in the representations observed between 100 and 200 ms. These early signals do not appear to drive behavior. Instead, suppression involves transforming these signals into an inverted format that downregulates the distractor's priority within a spatial priority map (Fig. 6). The timing of this transformation coincides with the emergence of target representations, suggesting that it may be driven by top-down feedback supporting target guidance rather than by feedforward attenuation alone. Notably, the later inverted singleton representations predicted faster reaction times, providing functional evidence that this transformation supports efficient behavior.

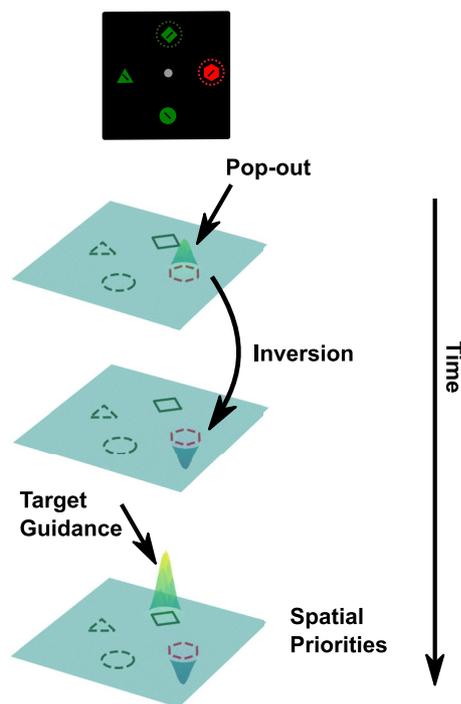


Figure 6. Proposed singleton suppression mechanism within the priority map framework. The process initiates with the registration of singleton distractor information. Subsequently, distractor representations undergo inversion, around the same time as the emergence of target representations. Both distractor and target representations are coded within a shared neural space but in an inverted manner. The representations within this neural space are then accessed to compute the ultimate spatial priorities that guide visual attention.

These findings bear directly on major theoretical accounts of singleton suppression. Both the signal suppression hypothesis (Sawaki and Luck, 2010; Gaspelin and Luck, 2018b) and the rapid disengagement hypothesis (Theeuwes, 2010) posit a transition from initial priority signals to suppression but differ in the proposed timing. Our results identify a representational transformation beginning ~ 200 ms. Although the early singleton representations precede this transformation, they were coded in a format distinct from both later singleton representations and target representations. At face value, this pattern could be interpreted as consistent with rapid disengagement hypothesis. However, linking decoding strength to behavior revealed that early singleton representations did not predict reaction times, whereas later representations did. This dissociation suggests that the early distractor representations are unlikely to reflect attentional capture. Instead, these signals may reflect sensory registration of singleton distractors that neither drives attentional capture nor suppression at that early stage of processing.

A limitation of the present study is the absence of eye-tracking data, which precludes directly testing whether early singleton representations reflect covert attentional capture. Nevertheless, several factors argue against this interpretation. First, attentional capture by singleton distractors is rare in similar paradigms, with reported rates typically below 10% (Gaspelin et al., 2017; Adams et al., 2023; Stilwell et al., 2023). Second, when capture does occur, saccade latencies cluster ~ 200 ms (Gaspelin et al., 2017; Adams et al., 2023), coinciding with the onset of the observed transformation. Thus, even if capture occurs, it overlaps temporally with the transformation rather than preceding it. These considerations suggest that the observed representational inversion is unlikely to reflect a purely reactive consequence of attentional capture. Importantly, recent modeling work suggests that the distinction between proactive and reactive suppression may be less fundamental than traditionally assumed. A single mechanism, preventing early distractor signals from triggering an eye movement, can account for both suppression and capture effects depending on relative signal timing (Zhang et al., 2025). Our findings align with this framework by demonstrating how early salience signals can be transformed into suppression-supporting representations.

The two-stage pattern we observed parallels interpretations of ERP components associated with singleton processing. An early positivity (Ppc; 80–150 ms) is thought to reflect initial saliency or sensory imbalance, whereas the later Pd (100–300 ms) has been linked to inhibitory processing (Sawaki and Luck, 2010; Barras and Kerzel, 2016; Gaspelin et al., 2023). Multivariate decoding allows these stages to be distinguished in representational space, providing evidence for a transformation that bridges early saliency signals and later suppression.

Finally, our findings speak to broader theories of spatial priority maps, which integrate goal-driven, experience-driven, and stimulus-driven signals to guide attention (Fecteau and Munoz, 2006; Awh et al., 2012; Luck et al., 2021). Using cross-condition generalization, we found evidence that a common priority representation supports both target enhancement and distractor suppression. However, this interpretation rests on assumptions about shared priority maps that remain debated (Bisley and Mirpour, 2019; Wolfe, 2021). It remains possible that multiple priority maps or representational subspaces coexist, enabling parallel implementation of enhancement and suppression. Disentangling these possibilities remains an important challenge for future work.

In conclusion, we provide evidence for a representational inversion mechanism underlying singleton distractor

suppression. Early salience signals are transformed into inverted representations that suppress distractor locations within a shared neural space, supporting efficient visual search.

References

- Adam KCS, Serences JT (2021) History modulates early sensory processing of salient distractors. *J Neurosci* 41:8007.
- Adams OJ, Ruthruff E, Gaspelin N (2023) Oculomotor suppression of abrupt onsets versus color singletons. *Atten Percept Psychophys* 85:613–633.
- Arita JT, Carlisle NB, Woodman GF (2012) Templates for rejection: configuring attention to ignore task-irrelevant features. *J Exp Psychol Hum Percept Perform* 38:580–584.
- Awh E, Belopolsky AV, Theeuwes J (2012) Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends Cogn Sci* 16:437–443.
- Bacon WF, Egeth HE (1994) Overriding stimulus-driven attentional capture. *Percept Psychophys* 55:485–496.
- Barras C, Kerzel D (2016) Active suppression of salient-but-irrelevant stimuli does not underlie resistance to visual interference. *Biol Psychol* 121:74–83.
- Bisley JW, Mirpour K (2019) The neural instantiation of a priority map. *Curr Opin Psychol* 29:108–112.
- Chang S, Egeth HE (2019) Enhancement and suppression flexibly guide attention. *Psychol Sci* 30:1724–1732.
- Cosman JD, Lowe KA, Zinke W, Woodman GF, Schall JD (2018) Prefrontal control of visual distraction. *Curr Biol* 28:414–420.e3.
- Fecteau JH, Munoz DP (2006) Salience, relevance, and firing: a priority map for target selection. *Trends Cogn Sci* 10:382–390.
- Forschack N, Andersen SK, Müller MM (2017) Global enhancement but local suppression in feature-based attention. *J Cogn Neurosci* 29:619–627.
- Gaspelin N, Leonard CJ, Luck SJ (2015) Direct evidence for active suppression of salient-but-irrelevant sensory inputs. *Psychol Sci* 26:1740–1750.
- Gaspelin N, Leonard CJ, Luck SJ (2017) Suppression of overt attentional capture by salient-but-irrelevant color singletons. *Atten Percept Psychophys* 79:45–62.
- Gaspelin N, et al. (2023) The distractor positivity component and the inhibition of distracting stimuli. *J Cogn Neurosci* 35:1693–1715.
- Gaspelin N, Ma X, Luck SJ (2025) Signal suppression 2.0: an updated account of attentional capture and suppression. *Psychon Bull Rev* 32:2648–2668.
- Gaspelin N, Luck SJ (2018a) Combined electrophysiological and behavioral evidence for the suppression of salient distractors. *J Cogn Neurosci* 30:1265–1280.
- Gaspelin N, Luck SJ (2018b) The role of inhibition in avoiding distraction by salient stimuli. *Trends Cogn Sci* 22:79–92.
- Geng JJ, Won B-Y, Carlisle NB (2019) Distractor ignoring: strategies, learning, and passive filtering. *Curr Dir Psychol Sci* 28:600–606.
- Ipata AE, Gee AL, Gottlieb J, Bisley JW, Goldberg ME (2006) LIP responses to a popout stimulus are reduced if it is overtly ignored. *Nat Neurosci* 9:1071–1076.
- Itti L, Koch C (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res* 40:1489–1506.
- Jonides J, Yantis S (1988) Uniqueness of abrupt visual onset in capturing attention. *Percept Psychophys* 43:346–354.
- Kerzel D, Burra N (2020) Capture by context elements, not attentional suppression of distractors, explains the PD with small search displays. *J Cogn Neurosci* 32:1170–1183.
- Kerzel D, Huynh Cong S (2023) The PD reflects selection of nontarget locations, not distractor suppression. *J Cogn Neurosci* 35:1478–1492.
- Klink PC, Teeuwen RRM, Lorteije JAM, Roelfsema PR (2023) Inversion of pop-out for a distracting feature dimension in monkey visual cortex. *Proc Natl Acad Sci U S A* 120:e2210839120.
- Liesefeld HR, Liesefeld AM, Müller HJ (2021) Attentional capture: an ameliorable side-effect of searching for salient targets. *Vis Cogn* 29:600–603.
- Liesefeld HR, et al. (2024) Terms of debate: consensus definitions to guide the scientific discourse on visual distraction. *Atten Percept Psychophys* 86:1445–1472.
- Liesefeld HR, Müller HJ (2019) Distractor handling via dimension weighting. *Curr Opin Psychol* 29:160–167.
- Liesefeld HR, Müller HJ (2021) Modulations of saliency signals at two hierarchical levels of priority computation revealed by spatial statistical distractor learning. *J Exp Psychol Gen* 150:710.
- Luck SJ, Gaspelin N, Folk CL, Remington RW, Theeuwes J (2021) Progress toward resolving the attentional capture debate. *Vis Cogn* 29:1–21.
- Miller J, Patterson TUI, Ulrich R (1998) Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology* 35:99–115.
- Oxner M, van Moorselaar D, Müller MM, Theeuwes J (2025) A flash in the pan? Distractor suppression cannot be inferred from the early lateralized positivity. *J Cogn Neurosci* 37:2351–2369.
- Reddy L, VanRullen R (2007) Spacing affects some but not all visual searches: implications for theories of attention and crowding. *J Vis* 7:3.1–17.
- Sapountzis P, Antoniadou A, Gregoriou GG (2025) Diverse neuronal activity patterns contribute to the control of distraction in the prefrontal and parietal cortex. *PLoS Biol* 23:e3003008.
- Sawaki R, Geng JJ, Luck SJ (2012) A common neural mechanism for preventing and terminating the allocation of attention. *J Neurosci* 32:10725.
- Sawaki R, Luck SJ (2010) Capture versus suppression of attention by salient singletons: electrophysiological evidence for an automatic attend-to-me signal. *Atten Percept Psychophys* 72:1455–1470.
- Stilwell BT, Egeth H, Gaspelin N (2022) Electrophysiological evidence for the suppression of highly salient distractors. *J Cogn Neurosci* 34:787–805.
- Stilwell BT, Adams OJ, Egeth HE, Gaspelin N (2023) The role of salience in the suppression of distracting stimuli. *Psychon Bull Rev* 30:2262–2271.
- Theeuwes J (2010) Top-down and bottom-up control of visual selection. *Acta Psychol* 135:77–99.
- Theeuwes J (2023) The attentional capture debate: when can we avoid salient distractors and when not? *J Cogn* 6:35.
- Theeuwes J (2025) Attentional capture and control. *Annu Rev Psychol* 76:251–273.
- Turatto M, Bonetti F, Pascucci D, Chelazzi L (2018) Desensitizing the attention system to distraction while idling: a new latent learning phenomenon in the visual attention domain. *J Exp Psychol Gen* 147:1827–1850.
- Ulrich R, Miller J (2001) Using the jackknife-based scoring method for measuring LRP onset effects in factorial designs. *Psychophysiology* 38:816–827.
- van Moorselaar D, Lampers E, Cordesius E, Slagter HA (2020) Neural mechanisms underlying expectation-dependent inhibition of distracting information. *Elife* 9:e61048.
- Vatterott DB, Vecera SP (2012) Experience-dependent attentional tuning of distractor rejection. *Psychon Bull Rev* 19:871–878.
- Wolfe JM (2021) Guided Search 6.0: an updated model of visual search. *Psychon Bull Rev* 28:1060–1092.
- Wolff MJ, Jochim J, Akyürek EG, Buschman TJ, Stokes MG (2020) Drifting codes within a stable coding scheme for working memory. *PLoS Biol* 18:e3000625.
- Won B-Y, Kosoyan M, Geng JJ (2019) Evidence for second-order singleton suppression based on probabilistic expectations. *J Exp Psychol Hum Percept Perform* 45:125–138.
- Won B-Y, Forloines M, Zhou Z, Geng JJ (2020) Changes in visual cortical processing attenuate singleton distraction during visual search. *Cortex* 132:309–321.
- Wöstmann M, et al. (2022) Ten simple rules to study distractor suppression. *Prog Neurobiol* 213:102269.
- Zhang H, Sellers J, Lee TG, Jonides J (2025) The temporal dynamics of visual attention. *J Exp Psychol Gen* 154:435–456.
- Zhang Z, Gaspelin N, Carlisle NB (2020) Probing early attention following negative and positive templates. *Atten Percept Psychophys* 82:1166–1175.
- Zhang Z, Sahatdjan R, Carlisle NB (2022) Benefits from negative templates in easy and difficult search depend on rapid distractor rejection and enhanced guidance. *Vision Res* 197:108031.
- Zhang Z, Carlisle NB (2023) Assessing recoding accounts of negative attentional templates using behavior and eye tracking. *J Exp Psychol Learn Mem Cogn* 49:509–532.