Perceptual distraction causes visual memory encoding intrusions

Blaire Dube[1] · Julie D. Golomb[1]

Abstract

Given the complexity of our visual environments, a number of mechanisms help us prioritize goal-consistent visual information. When searching for a friend in a crowd, for instance, visual working memory (VWM) maintains a representation of your target (i.e., your friend's shirt) so that attention can be subsequently guided toward target-matching features. In turn, attentional filters gate access to VWM to ensure that only the most relevant information is encoded and used to guide behavior. Distracting (i.e., unexpected/salient) information, however, can also capture your attention, disrupting search. In the current study we ask: does distraction also disrupt control over the VWM filter? Although the effect of distraction on search behavior is heavily studied, we know little about its consequences for VWM. Participants performed two consecutive visual search tasks on each trial. Stimulus color was irrelevant for both search tasks, but on trials where a salient distractor appeared on Search 1, we found evidence that the color associated with this distractor was incidentally encoded into VWM, resulting in memory-driven capture on Search 2. In two different experiments we observed slower responses on Search 2 when a non-target item matched the color of the salient distractor from Search 1; this effect was specific to the color associated with salient distraction and not induced by other non-target colors from the Search 1 display. We propose a novel Filter Disruption Theory: distraction disrupts the attentional filter that controls access to VWM, resulting in the encoding of irrelevant inputs at the time of capture.

Keywords Visual working memory · Visual search · Attentional capture

Introduction

Our visual environments contain too much detail to process at a given time. To compensate, mechanisms such as attention and visual working memory (VWM) help to constrain processing of these environments to only the aspects most relevant to a current behavioral goal. Effective behavior often relies on interactions between attention and VWM, which are entwined in many contexts (Chun et al., 2011; Dowd et al., 2017). Efficient visual search performance, for instance, is often supported by controlled, reciprocal interactions between attention and VWM. That is, the contents of VWM bias attention during visual search to facilitate target-finding and, in turn, attention plays a critical role in regulating what information is encoded in VWM.

A number of studies have demonstrated ways in which the contents of VWM can bias attention during visual search. During the search for a friend wearing a blue t-shirt on a busy street, for instance, VWM – our short-term storage system for visual information – maintains a representation of your target (known as an attentional template, i.e., your friend's blue t-shirt) (Carlisle et al., 2011; Woodman & Arita, 2011), so that attention can subsequently be guided towards visual inputs sharing its features (Bundesen, 1990; Desimone & Duncan, 1995). When the contents of VWM are relevant to the search task (e.g., searching for your friend's blue shirt), the VWM representation can be used to *intentionally* guide search. For example, an attentional template maintained in VWM supports the *facilitation* of target-matching stimuli in the environment (Carlisle et al., 2011; Woodman et al., 2007; Woodman & Arita, 2011). VWM resources can also be used in "visual marking" during search to support the *inhibition* of previously attended non-targets to prevent their re-selection (Al-Aidroos et al., 2012; Dube et al., 2016; Watson & Humphreys, 1997). VWM contents, however, are not always relevant to the current search; yet even under these circumstances, VWM can still unintentionally guide attention. Consider a task in which

✉ Blaire Dube
  dube25@osu.edu

[1] Department of Psychology, The Ohio State University, Columbus, OH 43210, USA

Springer

participants remember a color for a memory test, and a visual search task intervenes in which they locate a diamond among circles (i.e., Olivers et al., 2006). During search, if one of the circles is uniquely colored, attention will be captured by this "singleton distractor," and participants will be slower to find the target diamond. Critically, this reaction time (RT) slowing is *exacerbated* when the singleton's color matches the color concurrently maintained in memory, even if color is irrelevant. This effect is known as *memory-driven* capture (Olivers, 2009; Olivers et al., 2006).

In addition to VWM driving the guidance of attention, research has shown that goal-driven attention reciprocally plays a critical role in guiding VWM encoding. Specifically, given that VWM is severely capacity limited (Cowan, 2001), its effective use relies on attentional filtering, or the ability to effectively prioritize relevant – and *filter out* irrelevant – information in the environment. Thus, items (or spatial locations) that are prioritized for attention are more likely to be remembered (i.e., Schmidt et al., 2002). Moreover, performance in a VWM change-detection task has been shown to reflect an observer's ability to ignore irrelevant information in the initial memory array. Specifically, when cued to attend only to objects presented in a specified target color in a memory array, participants who were worse at attentional filtering (i.e., who encoded more memory-irrelevant items of the non-target color) exhibited poorer memory on the subsequent change-detection task (Vogel et al., 2005; Vogel & Machizawa, 2004). Using a similar task adapted for fMRI, McNab and Klingberg (2008) subsequently demonstrated that neural activity associated with the preparation to exclude distractors (i.e., "filter set activity") was negatively correlated with activity associated with VWM load. As such, attention is thought to serve as a VWM filter, restricting goal-irrelevant information from being unnecessarily encoded into capacity-limited VWM.

It is clear that effective attentional filtering is necessary for proficient VWM performance. What happens if this filter is disrupted, for example, under conditions of dynamic salient distraction? Distracting (i.e., unexpected and/or physically salient) information appearing during visual search can capture attention, compromising search performance. That is, if an ambulance suddenly appears in view during the search for your friend, rather than being guided toward target-matching visual inputs, spatial attention can be momentarily guided toward distracting information (i.e., the ambulance), increasing search RTs – a phenomenon generally known as *attentional capture*. A wealth of research has focused on the mechanics of attentional capture during visual search, demonstrating its consequences for attention (e.g., search RTs, accuracy, and eye movements; Bacon & Egeth, 1994; Bacon & Egeth, 1997; Folk et al., 2002; Folk et al., 1992; Theeuwes, 1991; Theeuwes, 1992; Theeuwes et al., 1998; Yantis & Jonides, 1984) and, more recently, perception (Chen et al., 2019).

Less explored is whether distraction affects other ongoing cognitive processes, such as what is encoded into VWM.

Here we propose a novel, previously untested consequence of distraction. We hypothesize that distraction not only disrupts the focus of attention for current perception (i.e., attracting spatial attention to the distractor location), but also disrupts the filter-gating VWM encoding such that goal-irrelevant features are encoded into VWM. The implication is not simply that a distractor's identity is encoded, but that completely irrelevant features (typically blocked by the attentional filter) also intrude into VWM if associated with the distractor. Thus, following attentional capture, *irrelevant* distractor features are incidentally encoded into memory, and once there, guide attention on future tasks.

To evaluate this hypothesis, we capitalize on memory-driven capture effects described above (Olivers, 2009; Olivers et al., 2006). In Experiment 1 (pre-registered), participants perform two consecutive visual search tasks on each trial. In the first search, they locate a target letter "T" among non-target letter "L"s, all presented within (irrelevant) colored squares. A distracting white border sometimes flashes briefly surrounding a non-target square, capturing attention. We hypothesize that, following capture, the features associated with this Search 1 salient distractor – including the task-irrelevant color – will be encoded into memory, thus impacting later search. In Search 2 participants locate a uniquely oriented landolt C stimulus. The Search 2 items are all white except one colored singleton; critically, its color sometimes matches a color from the Search 1 display, including the salient distractor (the color associated with the white border). We evaluate our prediction that the color associated with the Search 1 salient distractor will be encoded by assessing for memory-driven capture in Search 2 – that is, lengthened RTs on Search 2 trials when the singleton distractor matches the color of the salient Search 1 distractor, relative to other colors viewed in Search 1. Experiment 2 replicates Experiment 1, but with Search 2 stimuli that encourage feature detection mode (i.e., more heterogeneous search distractors) rather than singleton detection mode (i.e., more homogeneous search distractors in Experiment 1) (Bacon & Egeth, 1994; Lamy & Egeth, 2003).

Methods

Pre-registration

Participant recruitment techniques, target sample size, exclusion criteria, stimuli, task design, and procedure, and all data trimming routines and statistical analyses for Experiment 1 were preregistered on the Open Science Framework prior to data collection. Experiment 2 was conducted after completion of Experiment 1 using identical analyses.

Participants

Thirty-nine undergraduate students (19 female, 18 male, *M* age = 18.98 years) from The Ohio State University participated in Experiment 1 for partial course credit, and 40 Amazon Mechanical Turk workers (13 female, 27 male, *M* age = 37.95 years) participated in Experiment 2 for monetary compensation at a rate of $10/h. All participants reported having normal (or corrected-to-normal) visual acuity and color vision. We conducted a power analysis in advance of data collection based on typical effect sizes for memory-driven attentional capture effects (Olivers, 2009; Olivers et al., 2006), and estimated that 32 participants would be required to detect this effect of interest with 80% power. As pre-registered, we collected data from additional participants in anticipation of performance-based exclusions (i.e., > 20% response error on either Search 1 or Search 2). Based on these criteria, five participants were excluded from Experiment 1 and eight participants were excluded from Experiment 2, leaving final samples of 34 and 32 participants in Experiments 1 and 2, respectively. The study protocol was approved by The Ohio State University Behavioral and Social Sciences Institutional Review Board.

Apparatus

Stimuli in Experiment 1 were presented using PsychoPy (Peirce, 2007) on a 1,080 × 1,920 LCD monitor (calibrated using a Minolta CS-100 colorimeter; Minolta, Osaka, Japan) with a 240-Hz refresh rate. Viewing distance was fixed at 61 cm using a head and chin rest. Experiment 2 was presented online using PsychoPy and PsychoJS and was hosted on Pavlovia (pavlovia.org).

Stimuli and procedure

Stimuli and procedure for Experiments 1 and 2 were identical with the exception detailed below. Sizes are based on Experiment 1 and were approximated as closely as possible in Experiment 2 given the online testing environment. See Fig. 1 for a trial schematic. Each trial began with the presentation of a white fixation point (radius = .07°) in the center of the screen for 1.2 s. Search 1 then started with four colored squares appearing on the screen (size = 1.75 × 1.75° each, arranged in an invisible 2 × 2 grid centered on fixation with an eccentricity of 3°): three squares had the black letter "L" written in the center and one had the black letter "T" written in the center (text height = .5°). Letters could appear upright or inverted. The colors filling the squares were drawn for each trial from a 360° color wheel in CIE L*a*b* space (centered at L* = 70, a* = 20, b* = 38), selected with the constraint that the four colors used on a single trial were a minimum of 60° from each other on the wheel. All stimuli were viewed on a black background (8.2 cd/m^2). Participants searched for the target letter (the letter "T") and indicated whether it appeared upright or inverted by pressing "Z" for upright and "X" for inverted with their left hand; they were instructed to respond as quickly and accurately as possible. Color was irrelevant for the task, and participants were instructed to attend only to the letters within the squares. On 40% of trials, a salient distractor (a brief, sudden onset of a white border surrounding one of the non-target squares) appeared during Search 1. The distractor onset was .05 s after the onset of the Search 1 array and remained on screen for .1 s. As such, there were two Search 1 conditions: salient distractor present (240 trials) and salient distractor absent (360 trials). Search 1 terminated with the participant's response, or after 2 s if no response was recorded.

After a 1-s delay (during which only the fixation point appeared on the screen), Search 2 started with 8 landolt-cs (size = .88 × .88°) appearing on the screen arranged in a circle with 4.33° eccentricity. One of the landolt-cs (the target) had a gap on the right or left side. In Experiment 1, the other seven stimuli had gaps on the top or bottom; in Experiment 2, the seven non-target landolt-cs were rotated more heterogeneously such that their gaps could be at 0° (i.e., on the top), 45°, 135°, 180° (i.e., on the bottom), 22°, or 315°. In both experiments, participants located the target landolt-c and indicated which side the gap appeared on by pressing the right or left arrow key with their right hand and were instructed to do so as quickly and accurately as possible. On each trial, all of the Search 2 stimuli were colored white except for one of the non-target landolt-c's, which was a uniquely colored singleton. We manipulated the color of this singleton to create four Search 2 conditions: The singleton color could be an entirely novel color to that trial, match the color of a non-critical item from Search 1 (i.e., a square containing a non-target 'L'), match the color of the square that carried the target letter "T" from Search 1, or match the color associated with the salient distractor from Search 1 (i.e., the square framed by the salient white border on distractor present trials). As such, the four Search 2 conditions are referred to as: no match, non-critical match, target match, and salient distractor match. Search 2 terminated with the participant's response, or after 3 s if no response was recorded. Participants completed 600 trials. Experimental trials were divided equally among these four conditions for trials on which there was a salient distractor present in Search 1, resulting in 60 trials per Search 2 condition for the main analyses. For trials on which there was no salient distractor present in Search 1, Search 2 trials were divided equally among the no match, non-critical match, and target match singleton distractor conditions.

Analyses

Reaction times As preregistered, our main analyses compare mean RTs across conditions. We exclude full trials (data for
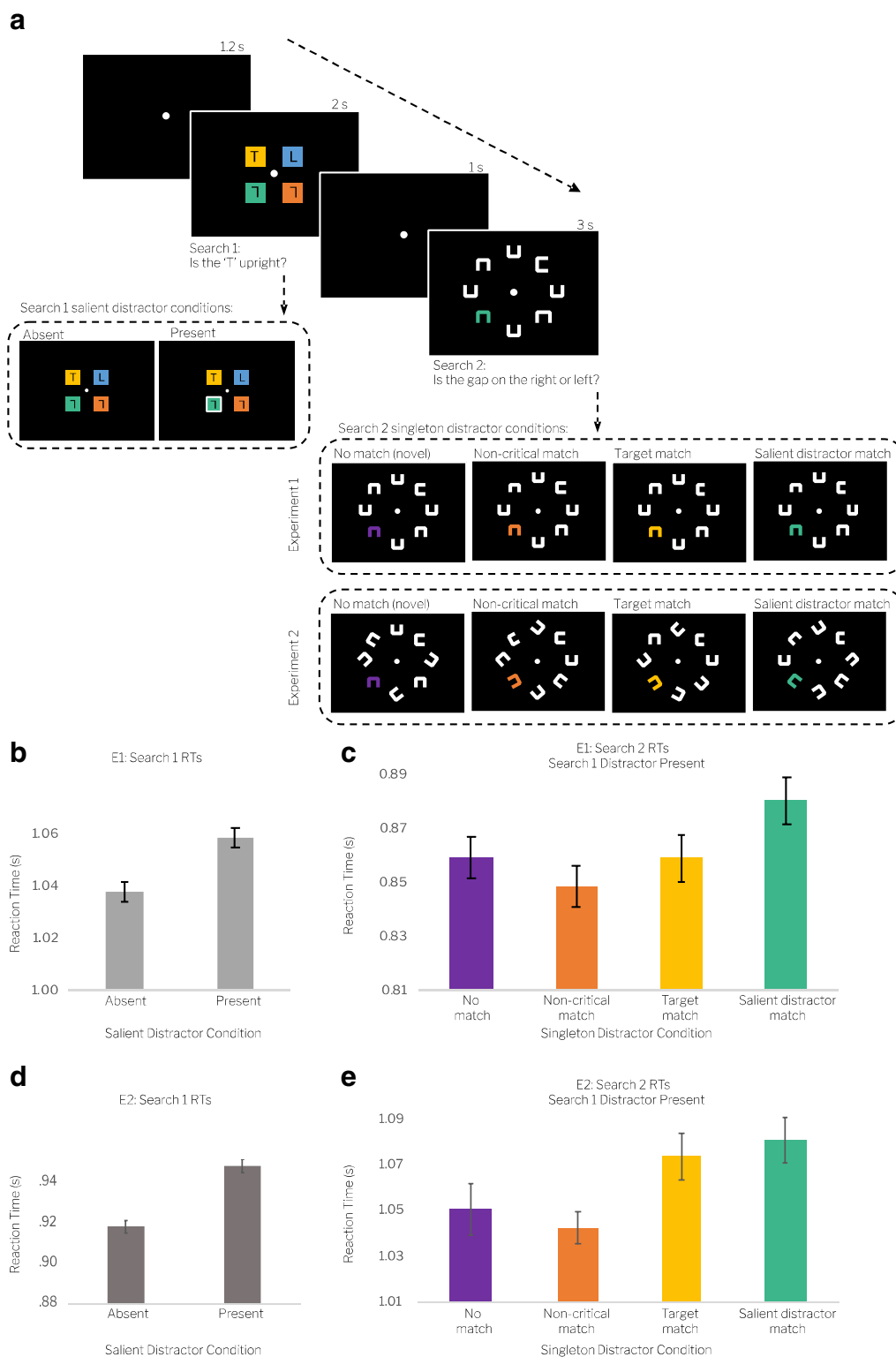
Fig. 1 **a** Trial schematic with condition breakdowns for Search 1 and Search 2 in both experiments. B-C, RT results for Experiment 1, during Search 1 (**b**) and Search 2, for trials on which there was a Search 1 distractor present (**c**). **d-e**, same for Experiment 2. In both experiments, Search 1 RTs were significantly slower when a salient distractor was present, indicating attentional capture. Search 2 RT was significantly slower for the key comparison: salient distractor match compared to the non-critical match, in both experiments. Error bars are adjusted within-subjects standard error (Morey, 2008), $N = 34$ (Experiment 1), $N = 32$ (Experiment 2). Note that, consistent with the goal of Experiment 2, Search 2 RTs are substantially slower in Experiment 2 relative to 1.

both Search 1 *and* 2) that contain extreme RT values (>2.5 standard deviations of the per participant, per condition mean) or incorrect responses, for either Search 1 or Search 2. Effect sizes (mean and 95% CI) are provided for all analyses.

For each experiment, we perform an initial paired-samples t-test on Search 1 RTs to compare the salient distractor-present and salient distractor-absent conditions to confirm that the salient Search 1 distractor effectively elicits attentional capture (i.e., slower RTs on distractor-present trials).

The primary analyses of interest are then carried out on Search 2 RTs, specifically comparing the four conditions for the Search 2 trials on which a salient distractor was presented in Search 1. Our preregistered analysis plan stated that we first perform a 1 × 4 repeated-measures analysis of variance (ANOVA), followed by four planned paired comparisons: (1) An initial check of no match versus non-critical match to assess whether simply viewing a color in Search 1 is sufficient to grant it privileged access to memory (i.e., lengthened RTs in the non-critical-match condition relative to the no-match condition would indicate memory-driven capture). (2) Our primary comparison of interest comparing salient distractor match versus non-critical match to assess for evidence of exacerbated, memory-driven capture in the critical condition. We use non-critical match as the comparison condition for salient distractor match as it was designed to control for expectation and novelty of the singleton colors in the critical Search 2 conditions. There is some precedent to predict that distractor predictability facilitates search (i.e., Kristjánsson & Driver, 2008; Shurygina et al., 2019; Vatterott & Vecera, 2012), and that novel colors might attract attention to a greater degree than a just-seen color (i.e., Horstmann, 2005), so this critical test for evaluating our hypothesis compares the effects of two just-seen non-target colors, with the difference being that one color was previously associated with the salient distractor. (3 and 4) As a secondary question of interest, we also evaluate memory-driven capture in the target match condition, comparing target match versus non-critical match (same rationale as above), and target match versus salient distractor match. Additional exploratory paired comparisons are reported in the Table 1 (caption).

Accuracy To confirm that there were no speed-accuracy tradeoffs we also performed exploratory (i.e., non-registered) analyses on accuracy data. There were no systematic differences in accuracy across conditions on either Search 1 or Search 2 (Table 1).

Results

Search 1 In both Experiment 1 and Experiment 2, Search 1 RTs (Fig. 1b and d) were significantly longer in the salient distractor-present condition relative to the salient distractor-absent condition, Experiment 1: $t(33)$ = -3.88, p < .001, d =

-.72 (CI$_{95\%}$ [-1.03, -.29]), Experiment 2: $t(31)$ = -6.63, p < .001, d = -1.17 (CI$_{95\%}$ [-1.69, -.71]), confirming the Search 1 salient distractor captured attention as intended.

Search 2 Figure 1c and e shows Search 2 RTs for the four conditions on which the Search 1 salient distractor was present. The 1 × 4 (singleton condition: no match, non-critical match, target match, and salient distractor match) repeated-measures ANOVA yielded a significant main effect of singleton type in both Experiment 1, $F(3, 99)$ = 2.61, p = .05, η_p^2 = .07 (CI$_{90\%}$ [.005, .14]), and Experiment 2, $F(3, 93)$ = 3.59, p = .02, η_p^2 = .10 (CI$_{90\%}$ [.01, .19]). The results of the four planned comparisons detailed in the analysis section are described below.

Initial test: *Mere visual exposure to a Search 1 item is not sufficient to grant it privileged access to memory*. Consistent with previous work (i.e., Olivers et al., 2006), RTs did not significantly differ between the no-match and non-critical-match conditions: Experiment 1, $t(33)$ = .96, p = .35, d = .16 (CI$_{95\%}$ [-.18, .50]); Experiment 2, $t(31)$ = .67, p = .51, d = .19 (CI$_{95\%}$ [-.23, .47]),

Main question: *Is the color associated with the Search 1 salient distractor item automatically encoded into VWM?* We found memory-driven capture induced by the salient distractor-match condition in both experiments, as evidenced by significantly slower Search 2 RTs compared to the non-critical-match condition (the control condition equated for visual exposure): Experiment 1, $t(33)$ = -2.72, p = .01, d = -.47 (CI$_{95\%}$ [-.82, -.11]); Experiment 2, $t(31)$ = -3.33, p = .002, d = -.59 (CI$_{95\%}$ [-.96, -.21]). This result suggests that the color associated with the salient Search 1 distractor was inadvertently encoded into memory, causing exacerbated (i.e., memory-driven) capture in the subsequent Search 2 when its color reappeared as the singleton. In Experiment 1, Search 2 stimuli were designed with more homogenous non-targets, which may have encouraged a singleton detection mode (Bacon & Egeth, 1994; Lamy & Egeth, 2003) and potentially amplified or interacted with this effect. Critically, in Experiment 2 we replicated this main finding when the Search 2 stimuli were adjusted to be more heterogeneous to encourage feature detection mode.

Secondary question: *What about the Search 1 target item?* When comparing Search 2 RTs between the target match and non-critical-match conditions, we found mixed results. In Experiment 1 we did not observe reliable evidence that the color associated with the Search 1 target interacted with attention in Search 2 to a greater degree than a non-critical Search 1 color, $t(33)$ = -.92, p = .36, d = -.16 (CI$_{95\%}$ [-.50, .18]). In Experiment 2, however, RTs *did* significantly differ between the non-critical-match and target-match conditions, $t(31)$ = -2.59, p = .01, d = -.46 (CI$_{95\%}$ [-.82, -.09]), suggesting that the color associated with the Search 1 target did interact with attention in Search 2. This disparity may be due to increased

Table 1 Summary of mean accuracy and RT (ms) data for Search 1 and Search 2 in Experiments 1 and 2 (E1, E2). Accuracy did not significantly differ between Search 1 conditions in either experiment (E1: $t(33) = -.8$, $p = .43$; E2: $t(31) = 1.21$, $p = .24$). Accuracy also did not differ between Search 2 conditions in either Experiment, neither when the search 1 salient distractor was absent (E1: $F(2,66) = .40$, $p = .67$; E2: $F(2,62) = .17$, $p = .84$) nor when it was present (E1: $F(3,99) = .23$, $p = .87$; E2: $F(3,93) = 1.34$, $p = .26$). Pre-registered RT comparisons are reported in the text. Additional exploratory comparisons: E1 no match vs. target match, $t(33) = .03$, $p = .98$, $d = .01$ (CI95% [-.33, .34]), E1 no-match vs. distractor match: $t(33) = -1.88$, $p = .07$, $d = -.32$ (CI95% [-.66, .03]), E2 no match vs. target match, $t(31) = -1.45$, $p = .16$, $d = -.26$ (CI95% [-.61, .10]), E2 no-match vs. distractor match: $t(31) = -1.91$, $p = .07$, $d = -.34$ (CI95% [-.69, .02])

	Search 1			Search 2		
	Salient Distractor Condition	Accuracy (SD)	RT (SD)	Singleton Distractor Condition	Accuracy (SD)	RT (SD)
E 1	Absent	.93 (.08)	1,038 (164)	No match	.97 (.04)	866 (164)
				Non-critical match	.97 (.04)	871 (157)
				Target match	.97 (.04)	870 (188)
	Present	.92 (.08)	1,059 (169)	No match	.97 (.03)	859 (160)
				Non-critical match	.97 (.04)	848 (177)
				Target match	.97 (.04)	859 (180)
				Salient distractor match	.97 (.04)	880 (177)
E 2	Absent	.96 (.04)	918 (137)	No match	.98 (.03)	1,051 (224)
				Non-critical match	.98 (.03)	1,046 (219)
				Target match	.98 (.02)	1,062 (226)
	Present	.96 (.04)	949 (142)	No match	.98 (.03)	1,051 (232)
				Non-critical match	.98 (.05)	1,042 (218)
				Target match	.98 (.03)	1,074 (251)
				Salient distractor match	.99 (.03)	1,081 (214)

power to detect differences in Experiment 2 afforded by the overall slower Search 2 RTs.

Comparison of target and salient distractor effects. When directly comparing RTs in the target match and salient distractor match conditions in Search 2, there was a nominal difference in Experiment 1, $t(33) = -1.96$, $p = .06$, $d = -.36$ (CI$_{95\%}$ [-.68, .01]), and no significant difference in Experiment 2, $t(31) = -.52$, $p = .61$, $d = -.09$ (CI$_{95\%}$ [-.44, .26]). As such, we do not find reliable evidence that the color associated with the distractor was encoded into memory to a significantly greater degree than the color associated with the Search 1 target. It is interesting to note that the target-associated intrusions into VWM were less consistent across the two experiments with more variable effect sizes, whereas the distractor-associated intrusions (our primary effect of interest) were replicated with large effect sizes across both experiments.

General discussion

In two Experiments we used a sequential search paradigm to examine whether irrelevant distractor features are incidentally encoded into VWM during attentional capture. We observed evidence for memory-driven capture in the second search when its display contained a non-target singleton that shared an irrelevant feature with the distractor from the first search.

That is, RTs in Search 2 were slower when the non-target singleton was the same color as that associated with the Search 1 salient distractor than when it was associated with a viewed but non-critical item from Search 1, despite color being irrelevant in both tasks. We suggest that, in addition to slowing RTs during the current search (Search 1), distraction also disrupts the filter that typically restricts irrelevant information (here color) from VWM encoding, causing the unnecessary storage of distractor features. These incidentally encoded distractor features can then drive attention during a subsequent visual search (Search 2).

We formulate this idea as the Filter Disruption Theory, which provides a theoretical framework to explain this novel consequence of distraction. Specifically, in addition to disrupting spatial attention, dynamic distraction disrupts the filter that gates access to VWM, resulting in the intrusion of distractor features in VWM. As a consequence, attention is unnecessarily guided towards distractor-matching elements of the environment – even those features that are completely irrelevant to the task – compromising subsequent visual search efficiency.

Evaluating the consequences of distraction for VWM during visual search is critical given how closely attention and VWM are entwined. While VWM supports the guidance of attention both towards and away from relevant (i.e., target-matching) and irrelevant (i.e., previously searched) visual inputs, effective use of this capability requires control over VWM contents via an

attentional filtering mechanism to regulate encoding. Our understanding of this attentional filter, however, has thus far been limited to tasks with an explicit memory component, such as tasks that require memory recall following intentional manipulations of attention (i.e., Dube et al., 2017; Emrich et al., 2017; Vogel et al., 2005), or tasks that evaluate how perceptual distractions that are related to memory representations disrupt memory performance (Kiyonaga & Egner, 2014). Here we evaluate the VWM filter in a novel context with a less explicit memory component, demonstrating filter disruption following abrupt-onset attentional capture in visual search. Despite no explicit role for VWM in the paradigm described here (i.e., no memory task or benefit of maintaining information), distraction results in encoding of irrelevant information into VWM, and this disruption of the attentional filter influences subsequent behavior.

The data presented here revealing a VWM filter disruption are particularly striking considering the design elements of the experiment. Specifically, color is a feature that is always optimal to ignore: It is irrelevant to Search 1 and there is no benefit to encoding any of the display colors, and in Search 2, a colored singleton is present on 100% of trials and always coincides with a non-target, so attending to it (a known non-target) incurs a *cost* to performance by delaying target identification. As such, the task was specifically designed to de-incentivize attending to or encoding color information (and even to desensitize participants to color singletons in Search 2). Our finding that the color associated with the Search 1 distractor was encoded into VWM highlights the disruptive effect of dynamic distraction on attentional control and its interactions with VWM encoding. That is, even with experimental conditions intended to strengthen observer control over task performance, dynamic distraction disrupted the filter controlling VWM encoding, causing the intrusion of color information associated with a distractor in Search 1 into VWM in a way that allowed it to drive attention in Search 2, incurring a performance cost. The fact that we replicated this result regardless of whether participants were in singleton selection mode (Experiment 1) or feature detection mode (Experiment 2) – and across both in-lab and online contexts – further strengthens this novel finding.

The primary focus of this study – and the analysis for which we had the strongest pre-registered predictions – was whether the salient distractor would intrude into VWM. We had less specific predictions about the *target* match Search 2 distractor condition. On one hand, work by Chen and colleagues (Chen & Wyble, 2015, 2016) on attribute amnesia suggests that an attended (but not directly relevant) attribute of a search target is not recalled during a surprise memory test and is therefore not encoded; yet more recent work by Harrison et al. (2020) demonstrates evidence that this attribute may still have a biasing effect on subsequent search behavior. We observed mixed evidence for subsequent guidance by the target-matching feature across our two experiments, consistent with the idea that

irrelevant features associated with a target item might sometimes – but not always – be encoded into VWM. Yet critically, the irrelevant features associated with the salient distractor were consistently encoded into VWM.

The present results – and the framework of the Filter Disruption Theory – raise a number of intriguing questions for study. One question is whether sudden-onset distraction disrupts filtering more generally, or only for items at the spatial focus of attention. If the former, disruption of the filter could have resulted in the incidental encoding of *all* irrelevant features on distractor-present trials (i.e., encoding color across the full display), but we did not observe exacerbated capture in the non-critical-match condition relative to no match. If the latter, the features that intrude into VWM may depend on how quickly control over the filter is regained following distraction and where spatial attention is at that time. If control over the filter is not reinstated until after spatial attention moves back to the target, this might explain the intrusion of irrelevant *target*-matching features into VWM on some trials.

In sum, attention and VWM are tightly linked and, given their strong reciprocal influence, a great deal of cognitive control is required in order to regulate how and when they interact. Here we demonstrate a circumstance in which dynamic distraction disrupts the otherwise carefully controlled interaction between attention and VWM. During a typical visual search, such as the search for a friend on a busy street, attention and VWM have distinct roles and interact in important ways: VWM guides attention towards target-matching representations (i.e., items that match the color of your friend's t-shirt) and monitors and maintains relevant non-targets (i.e., the positions of nearby vehicles), and control over attention ensures that only relevant information is encoded into VWM to guide subsequent behavior. We have long known that the sudden appearance of an ambulance during this search will capture spatial attention, increasing the time it takes to find your target. In line with our Filter Disruption Theory, we show for the first time – and then replicate – that this unexpected distractor can also momentarily disrupt control over whether, and how, attention and VWM are interacting. With the incidental encoding of ambulance features predicted by the Filter Disruption Theory, attentional biasing is disrupted, and fewer VWM resources are available to represent relevant subsequent visual inputs, such as nearby vehicles – which could pose potentially serious consequences for real-world behavior.

References

Al-Aidroos, N., Emrich, S. M., Ferber, S., & Pratt, J. (2012). Visual working memory supports the inhibition of previously processed information: Evidence from preview search. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 643–663. https://doi.org/10.1037/a0025707

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics*, *55*(5), 485–496. https://doi.org/10.3758/BF03205306

Bacon, W. J., & Egeth, H. E. (1997). Goal-directed guidance of attention: Evidence from conjunctive visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(4), 948–961. https://doi.org/10.1037/0096-1523.23.4.948

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*(4), 523–547. https://doi.org/10.1037/0033-295X.97.4.523

Carlisle, N. B., Arita, J. T., Pardo, D., & Woodman, G. F. (2011). Attentional templates in visual working memory. *The Journal of Neuroscience*, *31*(25), 9315–9322. https://doi.org/10.1523/JNEUROSCI.1097-11.2011

Chen, H., & Wyble, B. (2015). Amnesia for object attributes: Failure to report attended information that had just reached conscious awareness. *Psychological Science*, *26*(2), 203–210. https://doi.org/10.1177/0956797614560648

Chen, H., & Wyble, B. (2016). Attribute amnesia reflects a lack of memory consolidation for attended information. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(2), 225–234. https://doi.org/10.1037/xhp0000133

Chen, J., Leber, A. B., & Golomb, J. D. (2019). Attentional capture alters feature perception. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(11), 1443–1454. https://doi.org/10.1037/xhp0000681

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*(1), 73–101. https://doi.org/10.1146/annurev.psych.093008.100427

Cowan, N. (2001). The magical number 4 in short term memory. A reconsideration of storage capacity. *Behavioral and Brain Sciences*, *24*(4), 87–186. https://doi.org/10.1017/S0140525X01003922

Desimone, R., & Duncan, J. S. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Dowd, E. W., Pearson, J. M., & Egner, T. (2017). Decoding working memory content from attentional biases. *Psychonomic Bulletin and Review*, *24*(4), 1252–1260. https://doi.org/10.3758/s13423-016-1204-5

Dube, B., Basciano, A., Emrich, S. M., & Al-Aidroos, N. (2016). Visual working memory simultaneously guides facilitation and inhibition during visual search. *Attention, Perception, and Psychophysics*, *78*(5). https://doi.org/10.3758/s13414-016-1105-8

Dube, B., Emrich, S. M., & Al-Aidroos, N. (2017). More than a filter: Feature-based attention regulates the distribution of visual working memory resources. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(10). https://doi.org/10.1037/xhp0000428

Emrich, S. M., Lockhart, H. A., & Al-Aidroos, N. (2017). Attention mediates the flexible allocation of visual working memory resources. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(7), 1454–1465. https://doi.org/10.1037/xhp0000398

Folk, C. L., Leber, A. B., & Egeth, H. E. (2002). Made you blink! Contingent attentional capture produces a spatial blink. *Perception and Psychophysics*, *64*(5), 741–753. https://doi.org/10.3758/BF03194741

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1030–1044. https://doi.org/10.1037/0096-1523.18.4.1030

Harrison, G. W., Kang, M., & Wilson, D. (2020). Remembering more than you can say: Re-examining "Amnesia" of attended attributes. *Acta Psychologica*, *1*(343), 103265. https://doi.org/10.1016/j.actpsy.2021.103265

Horstmann, G. (2005). Attentional capture by an unannounced color singleton depends on expectation discrepancy. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(5), 1039–1060. https://doi.org/10.1037/0096-1523.31.5.1039

Kiyonaga, A., & Egner, T. (2014). The working memory stroop effect: When internal representations clash with external stimuli. *Psychological Science*, *25*(8), 1619–1629. https://doi.org/10.1177/0956797614536739

Kristjánsson, Á., & Driver, J. (2008). Priming in visual search: Separating the effects of target repetition, distractor repetition and role-reversal. *Vision Research*, *48*(10), 1217–1232. https://doi.org/10.1016/j.visres.2008.02.007

Lamy, D., & Egeth, H. E. (2003). Attentional capture in singleton-detection and feature-search modes. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(5), 1003–1020. https://doi.org/10.1037/0096-1523.29.5.1003

McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience*, *11*(1), 103–107. https://doi.org/10.1038/nn2024

Morey, R. D. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4*(2), 61-64.

Olivers, C. N. L. (2009). What drives memory-driven attentional capture? The effects of memory type, display type, and search type. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(October), 1275–1291. https://doi.org/10.1037/a0013896

Olivers, C. N. L., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1243–1265. https://doi.org/10.1037/0096-1523.32.5.1243

Peirce, J. W. (2007). PsychoPy-psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Schmidt, B. K., Vogel, E. K., Woodman, G. F., & Luck, S. J. (2002). Voluntary and automatic attentional control of visual working memory. *Perception & Psychophysics*, *64*(5), 754–763. https://doi.org/10.3758/Bf03194742

Shurygina, O., Kristjánsson, Á., Tudge, L., & Chetverikov, A. (2019). Expectations and perceptual priming in a visual search task: Evidence from eye movements and behavior. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(4), 489–499. https://doi.org/10.1037/xhp0000618

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics, 49*(1), 83–90.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, *51*(6), 599–606. https://doi.org/10.3758/BF03211656

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science*, *9*(5), 379–385. https://doi.org/10.1111/1467-9280.00071

Vatterott, D. B., & Vecera, S. P. (2012). Experience-dependent attentional tuning of distractor rejection. *Psychonomic Bulletin and Review*, *19*(5), 871–878. https://doi.org/10.3758/s13423-012-0280-4

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 748–751. https://doi.org/10.1038/nature02447

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067), 500–503. https://doi.org/10.1038/Nature04171

Watson, D. G., & Humphreys, G. W. (1997). Visual Marking: Prioritizing Selection for New Objects by Top-Down Attentional Inhibition of Old Objects. *Psychological Review*, *104*(1), 90-122. http://dx.doi.org/10.1037/0033-295X.104.1.90

Woodman, G. F., & Arita, J. T. (2011). Direct electrophysiological measurement of attentional templates in visual working memory. *Psychological Science*, *22*(2). https://doi.org/10.1038/jid.2014.371

Woodman, G. F., Luck, S. J., & Schall, J. D. (2007). The role of working memory representations in the control of attention. *Cerebral Cortex*, *17*, i118–i124. https://doi.org/10.1093/cercor/bhm065

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 601–621. https://doi.org/10.1037/0096-1523.10.5.601