



A new technique for estimating the probability of attentional capture

Taylor J. Rigsby¹ · Brad T. Stilwell¹ · Eric Ruthruff² · Nicholas Gaspelin¹

Accepted: 5 December 2022 / Published online: 9 January 2023
© The Psychonomic Society, Inc. 2023

Abstract

Latency-based metrics of attentional capture are limited: They indicate whether or not capture occurred, but they do not indicate how often capture occurred. The present study introduces a new technique for estimating the probability of capture. In a spatial cueing paradigm, participants searched for a target letter defined by color while attempting to ignore salient cues that were drawn in either a relevant or irrelevant color. The results demonstrated the typical contingent capture effect: larger cue validity effects from relevant cues than irrelevant cues. Importantly, using a novel analytical approach, we were able to estimate the probability that the salient cue captured attention. This approach revealed a surprisingly low probability of attentional capture in the spatial cueing paradigm. Relevant cues are thought to be one of the strongest attractors of attention, yet they were estimated to capture attention on only about 30% of trials. This new metric provides an index of capture strength that can be meaningfully compared across different experimental contexts, which was not possible until now.

Keywords Attention · Visual search · Attentional capture · Salience · Spatial cueing

It often seems as if salient objects, such as brightly colored advertisements or blinking traffic signals, have an automatic ability to attract our attention. This phenomenon has been documented in many research studies, but a unified theory of attention capture has remained elusive. On the one hand, stimulus-driven accounts propose that salient items have inherent power to attract attention, even when they are task irrelevant (Theeuwes, 1992; Yantis & Jonides, 1984). On the other hand, goal-driven accounts propose that salient objects do not automatically capture attention; instead, objects attract attention only when they possess the defining features of the target (Folk et al., 1992). Despite 3 decades of research, there is still disagreement as to which theory provides the best account (for a review, see Luck et al., 2021).

This prolonged debate may be due, at least in part, to limitations of the metrics used to assess attentional capture. Current metrics can indicate whether capture occurred, but do not indicate the *probability* that capture occurred. As a result, it is unclear how frequently attention is captured by salient distracting items. The current study aims to address

this shortcoming by introducing a simple new approach for estimating the probability of attentional capture from mean response times (RTs). As will be shown, this approach is relatively easy to implement and can yield valuable insights about the strength of attentional capture in a given task.

Common metrics of attentional capture

A variety of latency-based metrics have been developed to assess attentional capture. One common metric comes from the *spatial cueing paradigm* (Folk et al., 1992; Folk & Remington, 1998). As depicted in Fig. 1A, Folk and Remington (1998, Exp. 1) had participants search displays of letters for a target of a specific color (e.g., red) and quickly indicate its identity (X or =). Shortly before the search display appeared, a uniquely colored cue (a *color singleton*) appeared at a random location. If this cue captures attention, then there should be a *cue validity effect*: RTs should be faster when the cue appears at the target location (*valid* cue) than at a distractor location (*invalid* cue). Interestingly, the magnitude of the cue validity effect depended strongly on the match between the target defining feature and the salient cue. When the target was defined as the red item, red cues produced a cue validity effect, whereas green cues did not (Fig. 1B). Conversely, when the target was defined as a green item, green cues produced a cue validity effect, whereas red cues did not. This pattern was

✉ Taylor J. Rigsby
trigsby1@binghamton.edu

¹ Department of Psychology, State University of New York at Binghamton, P.O. Box 6000, Binghamton, NY 13902-6000, USA

² The University of New Mexico, Albuquerque, NM, USA

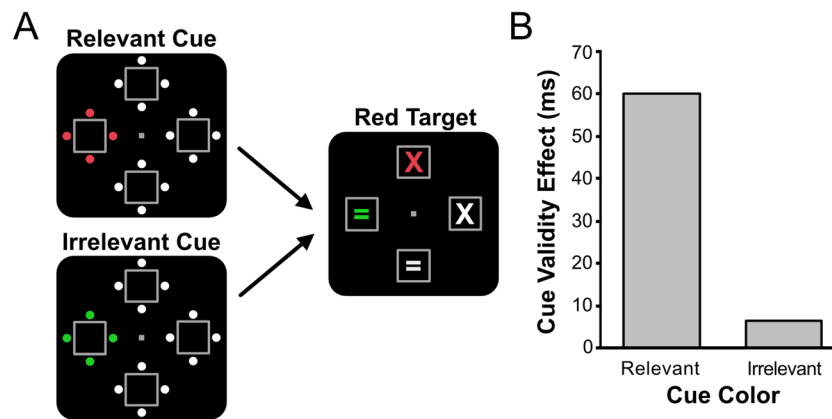


Fig. 1 A classic example of the spatial cueing paradigm (from Folk & Remington, 1998, Exp. 1, nonsingleton target condition). **A** Participants searched for a red target and attempted to ignore a salient cue that could

either match or mismatch the target color. **B** Cue validity effects were larger for salient items matching the target color (relevant) than salient items mismatching the target color (irrelevant). (Color figure online)

taken to indicate that salient items can capture attention only when task relevant (the *contingent capture account*; Folk et al., 1992). Since this original study, many studies have used cue validity effects to evaluate the strength of attentional capture by salient and task-relevant stimuli (e.g., Becker et al., 2010, 2013; Folk et al., 1992, 1994; Folk & Remington, 2008, 2015; Gaspelin et al., 2016; Irons et al., 2012; Lamy et al., 2004; Lien et al., 2010; Schönhammer et al., 2020; Zivony & Lamy, 2018).

Other analogous approaches have also been developed to assess attentional capture. For example, several studies have assessed whether the presence of a salient distractor slows RT during visual search for a target (e.g., Bacon & Egeth, 1994; Theeuwes, 1992). The underlying logic is that, if a salient distractor automatically captures attention, then the momentary misdirection of attention to the salient distractor should increase the time needed to locate the target. Other studies have assessed whether salient targets are located more efficiently than nonsalient targets as indexed by a reduction in search slope (e.g., Franconeri & Simons, 2003; Yantis & Egeth, 1999; Yantis & Jonides, 1984). Other studies have assessed response-time costs due to a semantic mismatch between the identity of the salient stimulus and the identity of the target stimulus (e.g., Maxwell et al., 2021; Remington & Folk, 2001; Zivony & Lamy, 2018).

In summary, a wide variety of latency-based metrics are used to evaluate whether salient distractors can automatically capture attention. A commonality amongst these metrics is that they all indicate a temporal cost incurred by misdirecting attention to the distractor item.

Limitations of current metrics of capture

Although widely used, latency-based metrics of attentional capture share an important limitation: They do not directly

indicate the *probability* that capture occurred. Instead, these metrics merely indicate a temporal cost associated with misdirecting attention to the distracting item. Often, the statistical significance of these latency-based effects is used to establish whether or not attentional capture occurred. However, this is problematic because attentional capture is likely not “all or none,” occurring on all trials or no trials. Instead, attentional capture is likely probabilistic in nature, occurring on some trials but not others. Assuming that visual objects compete for attention on some kind of priority map, salient distractors should win the competition for attention on some trials but not others (e.g., see discussions by Lamy, 2021; Leonard, 2021). Latency-based metrics provide no indication of the proportion of trials that salient distractors capture attention. This limitation of latency-based metrics has clearly impeded the development and testing of theories of attentional capture. Most theories of attentional capture specify whether or not capture occurs by certain kinds of stimuli, but do not specify how often it occurs or attempt to quantify how strongly certain factors, such as salience or task-relevance, modulate the strength of capture.

To illustrate, consider the contingent capture effect depicted in Fig. 1B (from Folk & Remington, 1998, Exp. 1), in which cue validity effects are larger for relevant cues (60 ms) than irrelevant cues (7 ms). This pattern of results was taken to suggest that relevant cues capture attention, whereas irrelevant cues do not. This interpretation is certainly justified. But it is unclear *how* likely the relevant cue is to capture attention. Top-down guidance might be so powerful that the lone object with the target-defining feature (e.g., the target color) captured attention on nearly 100% of trials. Yet it is also possible that relevant cue captured attention on only a small proportion of trials (e.g., 30%). Based on the cue validity effects alone, there is no way to know which is the case. Similarly, the small amount of capture with irrelevant cues (7 ms, n.s.) is consistent with 0% capture, but might also be consistent with, say, 15% capture.

A critic might argue that although latency-based capture effects do not indicate the absolute probability of capture, they at least tell us the *relative* probability of capture across studies or experimental conditions. That is, the magnitude of a latency-based capture effect could be used to infer the relative underlying “strength” of attentional capture. For example, if one condition produces a cue validity effect of 141 ms and another only 28 ms, one might be tempted to argue that capture was about five times more likely in the former. However, even this more limited claim is not safe. The above numbers were taken from Gaspelin et al. (2016, Exp. 7), in which the same exact cue stimulus produced a 141-ms cue validity effect under difficult search, but only 28 ms under easy search. Critically, because the easy and difficult conditions were randomly intermixed within blocks and therefore unknowable at the time of the cue display, they presumably shared the exact same probability of capture by the cue. So, rather than assuming that there was five times as much capture, the authors assumed equal probabilities of capture, but differing latency costs incurred during subsequent visual search. This finding and others suggest that one cannot even safely draw inferences about the relative probability of capture across studies or experimental conditions, because latency-based metrics are sensitive not only to the probability of capture but also to the cognitive processes that occur after attentional capture, such as rejection of the attended item (Geng & DiQuattro, 2010; Ruthruff et al., 2020).

In summary, latency-based metrics of attentional capture are limited in that they do not directly indicate the underlying probability of capture. As a result, we do not know how powerfully salient and/or task-relevant stimuli capture attention. This represents a major shortcoming in attentional capture research.

Estimating the probability of capture

The current study introduces a new approach to estimate the probability of capture from mean RTs. This approach could be used in any attentional capture paradigm provided that it manipulates both (a) set size, and (b) cue validity. For example, the spatial cueing paradigm introduced above could easily be adapted to use our approach.

The mathematical details of this approach are explained in the [Appendix](#), but the basic logic will be explained conceptually here. The current approach defines attentional capture as when a distracting item (such as a salient cue) directs attention to a location so that it is searched first. If this location contains the target (i.e., a valid trial), then capture should eliminate the need for visual search, and there should be no set size effect on mean RT. Whenever capture does not occur, however, the

normal search process takes place, and the normal set size effect should be obtained. Thus, the greater the probability of capture, the greater the reduction in the set size effect on valid trials. The approach is illustrated visually in [Fig. 2](#), which plots, for each set size, mean RT for valid trials on the y-axis against mean RT for invalid trials on the x-axis. This will result in two points on this plot, one for each set size. The slope of the line between these two points indicates the probability of capture.

To understand why, consider two extreme scenarios: 0% capture and 100% capture. If the salient cue never captures attention (i.e., 0% capture), then there should be no spatial cueing effect on RT. Thus, valid and invalid RTs should be equal, regardless of the set size. Hence, 0% capture should produce a line with a slope of 1. If capture instead occurs on every trial (i.e., 100% capture), then there should be no set size effect on valid RT, resulting in a horizontal line with a slope of zero. That is, on valid trials, attention should be immediately directed to the target location, eliminating the need for search. Invalid trials, however, would still require search after initial capture, resulting in a set size effect on invalid RT. The same logic can be extended to intermediate probabilities of capture in between 0 and 100%: the more often a salient object captures attention, the more it will reduce the set size effect on valid trials.

The probability of capture can be estimated by the slope of the line between the two set size conditions. Let (X_2, Y_2) represent Set Size 2 and (X_8, Y_8) represent Set Size 8, with X representing invalid RT and Y representing valid RT. The slope can be calculated using the same equation used to calculate slope in any Cartesian plane:

$$\text{slope} = \frac{Y_8 - Y_2}{X_8 - X_2}. \quad (1)$$

This can also be expressed as:

$$\text{slope} = \frac{\text{Set Size Effect}_{\text{Valid}}}{\text{Set Size Effect}_{\text{Invalid}}} \quad (2)$$

As demonstrated in the [Appendix](#), the probability of capture can then be estimated according to the following equation:

$$p(\text{capture}) = 1 - \text{slope}. \quad (3)$$

In summary, we operationalize attentional capture as the elimination of set size effects on valid trials. There are two important issues to consider about this approach. First, because the current metric measures capture as a reduction in the set size effect on valid trials relative to invalid trials, the search task needs to be difficult enough to produce a measurable set size effect on invalid trials. Second, the current metric does not distinguish between a discrete capture process that

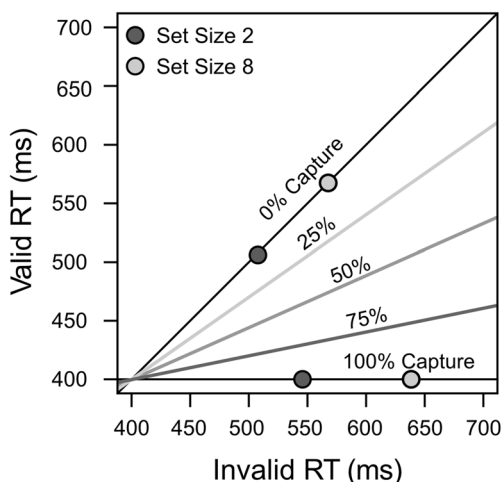


Fig. 2 A visual illustration of how the probability of capture is estimated. Each dot represents a different set size. As the probability of capture increases, the set size effect for valid cues decreases, and the slope becomes flatter

occurs fully on a subset of trials and one that is partial but occurs on all trials. In either case, our metric meaningfully estimates the strength of attentional capture and allows one to directly compare the strength of capture across studies and conditions. We will return to this issue in the General Discussion, after we first demonstrate the general success of the current approach.

Experiment 1

Experiment 1 used the technique outlined above to assess the probability of attentional capture in a spatial cueing paradigm. As shown in Fig. 3, participants searched for a target item of a specific color (e.g., pink) and reported its identity (T or L) via speeded button presses. Shortly before the search array appeared, a salient cue appeared at a randomly selected location

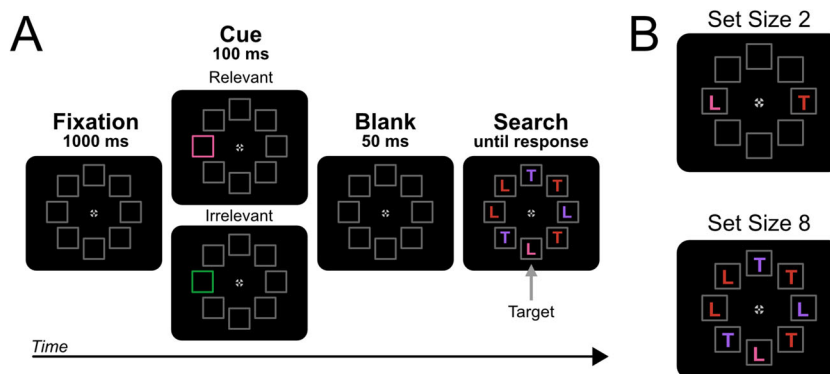


Fig. 3 The spatial cueing paradigm used in Experiment 1. **A** The progression of events in our spatial cueing paradigm. Participants searched for a target defined by color (e.g., pink) and reported its identity (T or L). This search array was preceded by a salient cue that

in either the target color (*relevant cue*) or a nontarget color (*irrelevant cue*). The search array set size of two and eight was varied between blocks.

Our technique to estimate the probability of capture involves calculating the slope, which is a ratio of two difference scores. Such ratios can be very noisy compared with a difference score, especially if the denominator can ever approach zero (a singularity point) due to measurement error (Franz, 2007). One potential solution to this problem is to accurately measure the denominator for each participant, so it is unlikely to approach zero. Accordingly, Experiment 1 used a *multisession approach* in which each participant completed many trials (3,264) over multiple sessions. This large number of trials should reduce measurement error (Rouder & Haaf, 2019) and is common in the psychophysical literature.

We expect to replicate the classic contingent capture effect of larger cue validity effects for relevant cues than irrelevant cues (e.g., Folk et al., 1992; Folk & Remington, 1998). There are two key questions regarding the probability of capture. The first question is how frequently capture will occur. If capture is powerful, it should occur on a large proportion of trials (e.g., 75%), whereas weak capture would occur on a relatively low proportion of trials (e.g., 25% of trials). The second question is how strongly top-down goals will influence the probability of capture.

Method

Participants

Sixteen participants from the State University of New York at Binghamton volunteered for course credit (11 women and five men; mean age = 25.0 years). All participants demonstrated normal color vision as assessed by an Ishihara test and self-reported normal or correct-to-normal visual acuity. All protocols were approved by a university ethics board and all participants gave informed consent.

could be either relevant (target color) or irrelevant (nontarget color). **B** Search display set size was two or eight items, manipulated between blocks. (Color figure online)

Apparatus

Stimulus presentation was accomplished using PsychToolbox (Brainard, 1997) on an Asus VG245H LCD monitor at a viewing distance of approximately 100 cm. A photosensor was used to measure the timing delay of the video system (12 ms), and this delay was subtracted from all latency values.

Stimuli

The spatial cueing task is depicted in Fig. 3A. Each search display contained eight gray placeholder squares ($1.9^\circ \times 1.9^\circ$) arranged in a notional circle around the center of the screen (3.9° in radius). Within the gray placeholders, letters (T or L) were drawn in Arial typeface (1.0° by 1.0°). These letters appeared in photometrically isoluminant colors: pink [36.0 cd/m^2 , $x = .485$, $y = .269$], green [36.0 cd/m^2 , $x = .270$, $y = .593$], red [36.0 cd/m^2 , $x = .571$, $y = .349$], purple [36.1 cd/m^2 , $x = .214$, $y = .120$], blue [36.2 cd/m^2 , $x = .189$, $y = .252$], yellow [36.0 cd/m^2 , $x = .405$, $y = .516$], and gray [33.8 cd/m^2 , $x = .313$, $y = .336$]. For the cue display, the color of one of the boxes (i.e., the cue) briefly changed to either pink or green. A gray fixation cross continuously appeared at the center of the screen. The fixation cross consisted of a gray circle (0.5° in diameter). Superimposed over the gray circle were two black rectangles in a cross (each 0.5° by 0.1°) with a gray dot in the center (0.1° in diameter).

Design

Each participant was assigned one target color (green or pink) for all their experimental sessions; target color was counterbalanced across participants. Nontarget letter colors were selected to ensure a relatively difficult visual search, which is necessary to produce a substantial set size effect, which is critical for accurately estimating the probability of capture. In the pink target version, the distractor colors were red and purple. In the green target version, the distractor colors were blue and yellow. In each search display, one letter was drawn in the target color and the remaining items were drawn in distractor colors.

The cue displays consisted of seven gray squares and one colored square (see Fig. 3A). The cue color could be either relevant or irrelevant, which varied randomly from trial to trial. Relevant cues were rendered in the target color and irrelevant cues were rendered in the unassigned target color. This maximized the distance in color space between the relevant and irrelevant cue colors.

We used two different search display set sizes, which were varied by block. Although our approach could in principle be used with more than two set size conditions (e.g., two, four, and eight items), our approach requires many trials per condition to obtain precise slope estimates for each participant.

Using only two set-size conditions allowed us to have more trials per condition. Additionally, we chose to block set size to ensure that the cues were task irrelevant. When Set Size 2 is blocked, participants know that the target will always appear in either the left or right location; a cue appearing at one of these two locations provides no information about the upcoming target location, so there is no incentive to attend it. If Set Sizes 2 and 8 were intermixed, however, a cue in the left or right position would inform the participant that the set size is likely two and that the target is likely to appear on the left or right. In other words, the cue would become informative, giving participants a reason to attend to it. This kind of voluntary attentional allocation to the cue would inflate the amount of capture. An alternative approach to mixing set sizes would be to allow the salient cue to appear at any of the eight locations (even at Set Size 2). However, this design has the major drawback that many of the trials would be unusable (i.e., the cue would appear at a nonsearch location on $6/8^{\text{th}}$ of trials), impairing our ability to accurately estimate capture. We therefore elected to use the very simple solution of blocking set size, which solves both of the issues with intermixing mentioned above.

In Set Size 8 blocks, letters appeared at all eight locations. One location contained a letter in the target color. The colors of the remaining distractors were selected at random, with the restriction that there were always four letters in one color (e.g., red), and three letters in the other color (e.g., purple). The location and identity of the target letter were selected at random on each trial. The identities of the distractor letters (T or L) were also selected at random with the restriction that a letter could not repeat more than four times in any given set size 8 display. In set size 2 blocks, letters appeared only in the left and right positions on the horizontal midline. One letter was drawn in the target color, and the other letter was drawn in a randomly selected distractor color.

The location of the cue was selected at random and was nonpredictive of the upcoming target location. Thus, in Set Size 2 blocks, the cue could appear with equal probability at either the left or right position on the horizontal midline. The cue appeared at the target location on 50% of trials (*valid cue*) and appeared at the nontarget location on 50% of trials (*invalid cue*). In Set Size 8 blocks, the cue appeared with equal probability at all eight locations and therefore had a 12.5% chance of being valid. The only exception to this was the fourth session, as will be explained below.

Procedure

Participants were instructed to maintain central fixation throughout the experiment. Each trial began with the fixation screen for 1,000 ms, followed by the cue display for 100 ms. Next, the fixation display appeared for 50 ms. Then, the search array appeared for 3,000 ms or until response.

Participants made speeded responses on a gamepad to the identity of the target letter (L or T) using the left and right shoulder buttons, respectively. If participants did not respond within 3,000 ms, a low beep (200 Hz) sounded, and the message “Too Slow!” appeared for 300 ms. If participants responded incorrectly, the screen went blank, and a low beep (200 Hz) sounded for 300 ms.

Each participant completed four sessions of 816 trials, resulting in 3,264 trials per participant. The first three sessions used a nonpredictive cue (2,304 trials) and the final fourth session used a predictive cue as a control condition (816 trials). Each individual session was separated by at least 1 day and took approximately 1 hour to complete. Sessions were divided into 14 blocks and separated in two phases, one for each set size. The Set Size 2 phase consisted of a practice block (24 trials) followed by four regular blocks (64 trials each; 256 total trials). The Set Size 8 phase consisted of a practice block (24 trials) followed by eight regular blocks (64 trials each; 512 total trials). We included more blocks of the Set Size 8 condition because valid trials were much rarer in this condition (12.5%) than the Set Size 2 condition (50%) because the cues were nonpredictive (i.e., $1/n^{\text{th}}$ of trials were valid, where n is the set size).

The order of the set size phases was counterbalanced across participants. That is, half of participants first completed Set Size 2 blocks and then set size 8 blocks. The other half completed the set size phases in the reverse order. At the end of each block, participants received feedback on mean RT and accuracy.

Session 4 (predictive cue control condition)

The fourth session was designed to test the assumption that our set size manipulation selectively influences the visual search stage, by verifying that 100% capture produces a slope of zero. It is conceivable that it would not. For example, mean RT could differ between set sizes on valid trials even with 100% capture due to decision noise (Palmer, 1995) or visual crowding (Whitney & Levi, 2011). If such extraneous influences of set size are substantial, it would suggest the need to correct the capture probability estimates. The fourth session therefore used a cue designed to maximize attentional capture (to near 100%). All methods were identical to the Sessions 1–3 except for two changes. First, we eliminated the irrelevant-color cues, so that all cues were relevant-colored. Second, the cue always appeared at the future target location (100% predictive). Participants were specifically informed of this manipulation and instructed that they should use the cue to find the target letter. This cue should strongly attract attention because participants could use it to effectively bypass visual search. If valid RT is approximately equal at both set sizes in this display, it would confirm that our probability estimation is working as intended without need for any correction.

Salience verification

Saliency maps were used to independently verify the salience of the color singletons in the cue displays (Chang et al., 2021; Stilwell et al., 2022; Stilwell & Gaspelin, 2021). All 16 potential cue displays (8 locations \times 2 potential cue colors) were analyzed in the Image Signature Toolbox (Hou et al., 2012) in MATLAB to generate saliency maps. We selected this toolbox because it has previously been shown to perform similar to human observers (Kotseruba et al., 2020). The default parameters were used, with the exception of the `mapWidth()` parameter, which was adjusted to accommodate the image resolution (1,920 \times 1,080). This resulted in a saliency map for each potential cue display (see Fig. 4). For each saliency map, a circular region of interest (2.84° in diameter) was defined around each of the eight cue positions. The mean saliency score at each location was calculated by averaging the pixels in the interest area.

We calculated two metrics of salience from the saliency maps. First, we calculated a global saliency index (GSI), which was the saliency rating of the color singleton cue minus the average of the uncued locations (i.e., gray placeholders). This difference score was then normalized by dividing it by the sum of the saliency scores of all locations. Values of this saliency score can range from -1 (indicating the cue was *less* salient than average uncued location) to 1 (indicating the cue was *more* salient than average uncued location). The average GSI across all images was 0.92, indicating that the cue was indeed highly salient. In addition, we calculated the *singleton win rate* as the proportion of trials in which the singleton cue was identified as the most salient object in the display. The singleton win rate was 100%, indicating that the cue was always the most salient item. Altogether, this analysis verifies that our cues were highly salient.

Data analysis

The data from practice blocks and the first trial of each experimental block were excluded from analysis. RT cutoffs were

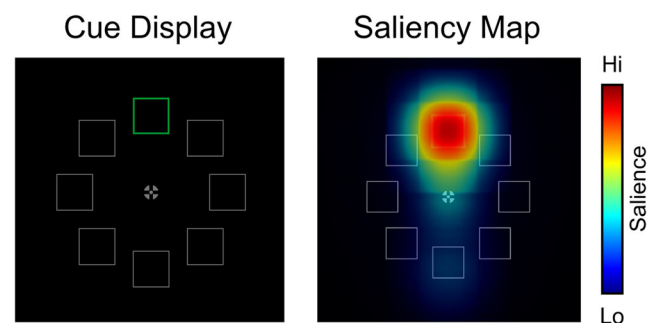


Fig. 4 An example cue display used in the present experiments (left) and the corresponding saliency map generated with the Image Signature Toolbox (right). As can be seen, the singleton cue was rated as highly salient. (Color figure online)

set for each participant at 2.5 standard deviations above and below their personal means for each set size. These cutoffs eliminated 2.6% of trials. Additionally, inaccurate responses were excluded from RT analyses (4.0% of trials). Cue validity effects were calculated for each condition by subtracting valid RTs from invalid RTs. Effect sizes were calculated utilizing Cohen's d_z and all partial eta squared values were adjusted for positive bias (Lakens, 2013; Mordkoff, 2019).

Calculating probability of capture

The probability of capture was calculated according to Equation 3 (i.e., $1 - slope$). Importantly, there are two potential methods of calculating the slope. One method is to calculate a slope for each participant by using their mean RTs in each condition. Then, these slopes could be averaged to form a grand average slope which could then be used to estimate the probability of capture (i.e., $1 - slope$). An alternative method is to first calculate grand average mean RT across participants for each condition. Then, grand average mean RTs can be used to calculate a single slope (e.g., as shown in the Fig. 2) used to determine the probability of capture. With additive metrics, such as difference scores, these two methods would produce the same result. But these two methods will not produce exactly the same estimates in the current case because slopes are a ratio that involves division (see, e.g., Simpson's paradox; Kievit et al., 2013). Because the denominator of the slope ratio (*Set Size Effect_{Invalid}*) could easily approach zero (a singularity point) for any individual due to measurement error, averaging across individual participant slopes is untenable (e.g., Franz, 2007). Slope estimates based on grand average mean RTs will more accurately represent the population mean, but do not provide an estimate of sample variability need for statistical tests.

To solve this problem, we used the jackknife technique (J. Miller et al., 1998; R. G. Miller, 1974; Ulrich & Miller, 2001), which is a bootstrapping-based approach that more accurately estimates the standard error in many situations. Jackknifing is widely used in studies that measure the latency of event-related potentials, which can have a similar problem of unstable data for individual participants. Jackknifing is also well-suited for estimating the standard error of ratio scores, which can be noisy with low numbers of trials (R. G. Miller, 1974; Oranje, 2006).

To implement jackknifing, the mean RT data for each condition were subsampled N times, each with $N-1$ participants; that is, each sample removed one of the participants. Then, these jackknifed mean RTs (N of them) were used to compute a slope with each participant removed. Importantly, the grand average of the jackknifed probability of capture estimates produces a value that is identical to the probability of capture based upon the grand average mean RT. The jackknifed means were used to compute standard errors used in the

statistical analyses (e.g., t tests). The resulting t values were then corrected by dividing them by the degrees of freedom (Ulrich & Miller, 2001). For those interested, the online data show exactly how we implemented the jackknifing technique. We will revisit the topic of jackknifing and its benefits in a dedicated section of the General Discussion.

Results

Response times

Table 1 depicts mean RTs and cue validity effects as a function of cue relevance, set size, and cue validity. As can be seen, we found the typical contingent capture effect of larger cue validity effects for relevant cues than irrelevant cues.

A three-way repeated measures analysis of variance (ANOVA) was conducted on mean RT with the factors set size (2 vs. 8), cue validity (valid vs. invalid), and cue relevance (relevant vs. irrelevant color).

There was a main effect of set size, $F(1, 15) = 100.94$, $p < .001$, $adj. \eta_p^2 = .86$, indicating that mean RTs were faster at set size 2 than set size 8. There was also a significant main effect of cue validity, $F(1, 15) = 19.34$, $p < .001$, $adj. \eta_p^2 = .53$, indicating that RTs were faster for valid trials than invalid trials. There was a main effect of cue relevance $F(1, 15) = 53.43$, $p < .001$, $adj. \eta_p^2 = .77$, indicating that RTs were faster for relevant cues than irrelevant cues.

According to contingent capture account, cue validity effects should be larger for relevant cues than irrelevant cues (Folk et al., 1992). The interaction of cue validity and cue relevance was significant, $F(1, 15) = 141.78$, $p < .001$, $adj. \eta_p^2 = .90$, indicating larger cue validity effects for relevant cues than for irrelevant cues. We had no other predictions about the other two-way interactions in the ANOVA. The interaction of set size and cue validity was significant,

Table 1 Mean response time (in ms) by cue type, set size, and cue validity

Experiment 1						
Cue Type	Set Size 2			Set Size 8		
	Invalid	Valid	CVE	Invalid	Valid	CVE
Relevant	507 (66)	483 (65)	24	573 (64)	531 (58)	42
Irrelevant	499 (72)	501 (65)	-2	569 (63)	568 (67)	1
Experiment 2						
Cue Type	Set Size 2			Set Size 8		
	Invalid	Valid	CVE	Invalid	Valid	CVE
Relevant	587 (67)	546 (57)	41	642 (67)	581 (65)	61
Irrelevant	579 (69)	567 (62)	12	637 (67)	618 (63)	19

Cue validity effects (CVE) were calculated as invalid minus valid for each respective set size by cue condition. Standard deviations are provided within parentheses

$F(1, 15) = 8.40, p = .011, \text{adj. } \eta_p^2 = .32$, indicating larger cue validity effects at Set Size 8 than Set Size 2. The interaction of set size and cue relevance was also significant, $F(1, 15) = 22.87, p < .001, \text{adj. } \eta_p^2 = .58$, indicating that set size effects (Set Size 8 minus Set Size 2) were larger for relevant cues than irrelevant cues. Finally, the three-way interaction of set size, cue validity, and cue relevance was significant, $F(1, 15) = 10.48, p = .006, \text{adj. } \eta_p^2 = .37$, indicating that the two-way interaction of cue validity and cue relevance (i.e., the contingent capture effect) to be greater at Set Size 8 than at Set Size 2.

Preplanned one-sample t -tests were conducted to assess the significance of each cue validity effect against zero for each cue type and cue relevance. For irrelevant cues, cue validity effects were not significantly different from zero at Set Size 2 (–2 ms), $t(15) = 0.50, p = .623, d_z = 0.13$, or Set Size 8 (1 ms), $t(15) = 0.21, p = .84, d_z = 0.06$. For relevant cues, however, cue validity effects were significantly greater than zero at both Set Size 2 (24 ms), $t(15) = 6.89, p < .001, d_z = 1.72$, and Set Size 8 (42 ms), $t(15) = 8.73, p < .001, d_z = 2.18$.

Probability of capture

The probability of capture was estimated as described above. Figure 5A (relevant cues) and 5B (irrelevant cues) plot mean RTs for valid and invalid trials at each set size. The resulting two coordinates formed a line in Cartesian space, whose slope was used to estimate the probability of attentional capture (i.e., $1 - \text{slope}$). For ease of comparison, Fig. 5C depicts the resulting probability of capture for each cue condition as a bar plot.

A paired-sample t test was performed to compare the probabilities of capture between the two cue types. To be clear, all t tests were computed on the jackknifed data and were therefore corrected (Ulrich & Miller, 2001). The probability of capture by relevant cues (28.5%) was significantly higher than that for

irrelevant cues (4.5%), $t(15) = 3.22, p = .006, d_z = 0.80$. Preplanned one-sample t tests compared the probability of capture for each cue type (relevant vs. irrelevant) against zero. The probability of capture was significantly greater than zero for relevant cues, $t(15) = 3.91, p = .001, d_z = 0.98$, but not for irrelevant cues, $t(15) = 0.78, p = .447, d_z = .20$.

Session 4: 100% predictive cues

An important assumption of the probability of capture estimate technique is that 100% capture should eliminate set size effects on valid trials. This would effectively produce a slope of zero (see Fig. 2). To test the assumption, the final session used cues that were 100% valid and were always presented in the target color (i.e., were relevantly featured). Participants were informed of this and told to use these cues to find the target. This should yield maximally strong attentional allocation by the cues (i.e., near 100% capture). The results confirmed the assumption that the slope would be near zero for 100% capture: mean RTs were nearly equal at Set Size 2 (449 ms) and Set Size 8 (450 ms), $t(15) = 0.29, p = .78, d_z = 0.07, \text{BF}_{01} = 3.77$. This suggests that, in the current study, set size effects were primarily due to differences in visual search time (which can be essentially eliminated with a maximally strong cue), rather than to decision noise or visual crowding.

Error rates

Error rates were generally quite low (3.7%). Although there were no key predictions, the same analyses on manual RT were repeated for error rates for the sake of completeness. There was a main effect of cue validity, $F(1, 15) = 6.88, p = .019, \text{adj. } \eta_p^2 = .27$, indicating that error rates were higher for invalid cues than valid cues. The interaction between cue validity and cue relevance was significant, $F(1, 15) = 6.47, p =$

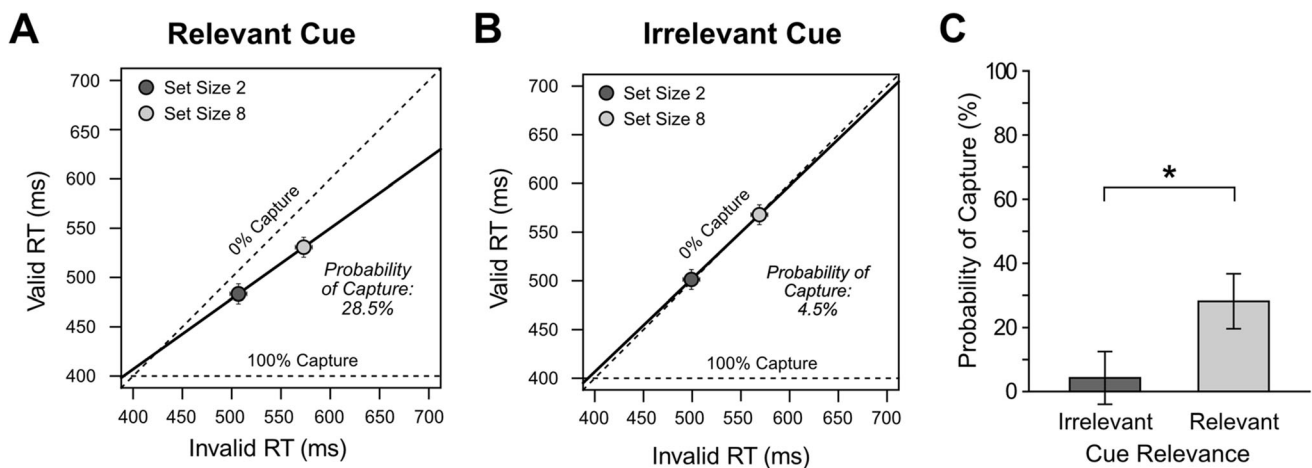


Fig. 5 Probability of capture estimates for Experiment 1. Response time on valid vs. invalid trials for (A) relevant cues (target color) and (B) irrelevant cues (nontarget color). C A comparison of the estimated

probability of capture for relevant cues and irrelevant cues. Error bars represent 95% within-subject confidence intervals

.023, *adj. η_p^2* = .26, reflecting a larger cue validity effect on error rates for relevant cues than irrelevant cues. All other main effects and interactions were nonsignificant ($ps > .10$).

Discussion

Experiment 1 estimated the probability of capture in the spatial cueing paradigm that has been classically used to study contingent capture (Folk et al., 1992). Participants searched for a color target while attempting to ignore salient color cues that either matched (relevant) or mismatched the target color (irrelevant). These cues were nonpredictive of the target location and therefore should have been ignored. The RT results were consistent with many previous contingent capture studies wherein relevant cues produced large cue validity effects and irrelevant cues did not (Folk et al., 1992; Folk & Remington, 1998; Lien et al., 2010).

Importantly, we used our new technique to estimate the underlying probability of capture for each cue type. Interestingly, the probability of capture estimate also showed a contingent capture effect: irrelevant cues produced a low probability of capture (4.5%) that did not significantly differ from zero, whereas relevant cues produced a higher probability of capture (28.5%). That being said, neither cue approached maximal capture and the estimates of capture were remarkably low. This occurred even though a control condition confirmed that 100% capture is possible when the cues reliably predicted the target location.

Experiment 2

Experiment 1 estimated the probability of capture using a multisession approach in which each participant completed many trials over multiple sessions. Experiment 2 investigated whether it is also feasible to do so with the more common and practical approach of having each participant complete a single session.

To equate the overall number of trials with Experiment 1, Experiment 2 used a correspondingly larger sample size ($N = 48$). As previously explained, the probability of capture estimate is a ratio score, which can be extremely noisy if the denominator ever approaches zero (a singularity point) due to measurement error. This problem is only exacerbated further by using single-session participants. To address this problem, we again used jackknifing to compute the probability of capture, as in Experiment 1. In jackknifing, means are based on the entire sample minus one participant, so the denominator should not approach zero.

The key questions are identical to Experiment 1. First, we will assess whether the overall probability of capture is lower than previously assumed, replicating Experiment 1. Second,

we will assess how strongly task-relevance influences the probability of attentional capture.

Method

Participants

A new sample of 48 participants from the State University of New York at Binghamton participated for course credit (26 women, 21 men, and one nonbinary; mean age = 19.0 years). This sample size was determined a priori to equate the total number of trials with Experiment 1 (i.e., 16 participants \times 3 nonpredictive sessions = 48 total sessions). One participant was replaced due to having abnormally slow mean RTs (>3.5 *SDs* from the group mean).

Stimuli, design, and procedure

All experimental procedures were identical to Experiment 1, except that each participant completed only a single experimental session with nonpredictive cues. The use of a single session per participant reduced the overall number of individual trials (816 trials).

Data analysis

The trimming procedures and the series of analyses were the same as those from Experiment 1. The data from practice blocks and the first trial of each experimental block were excluded from analysis. RT cutoffs were again set for each participant at 2.5 standard deviations above and below their personal means for each set size. These cutoffs eliminated 2.5% of trials. Additionally, inaccurate responses were excluded from RT analyses, eliminating 3.5% of trials.

Results

Response time

Table 1 lists mean RTs and cue validity effects for each experimental condition. As can be seen, Experiment 2 also found larger cue validity effects for relevant cues than irrelevant cues as predicted by the contingent capture account (Folk et al., 1992).

A three-way repeated-measures ANOVA was conducted on mean RTs with the factors set size, cue validity, and cue relevance. Overall, the results closely matched those of Experiment 1. Again, all three main effects were significant. There was a main effect of set size, $F(1, 47) = 53.79$, $p < .001$, *adj. η_p^2* = .52, indicating that mean RTs were faster at Set Size 2 than Set Size 8. The main effect of cue validity was also significant, $F(1, 47) = 128.04$, $p < .001$, *adj. η_p^2* = .73, indicating that RTs were faster for valid trials than invalid trials.

There was a main effect of cue relevance, $F(1, 47) = 33.25, p < .001, \text{adj. } \eta_p^2 = .40$, indicating that RTs for relevant cues were significantly faster than those for irrelevant cues.

The interaction of cue validity and cue relevance was significant, $F(1, 47) = 80.69, p < .001, \text{adj. } \eta_p^2 = .62$, indicating that cue validity effects were larger for relevant cues than irrelevant cues. No key predictions were made for the other interactions in this ANOVA. The interaction of set size and cue validity was significant, $F(1, 47) = 9.97, p = .003, \text{adj. } \eta_p^2 = .16$, indicating larger cue validity effects for Set Size 8 than Set Size 2. The interaction of set size and cue relevance was significant, $F(1, 47) = 9.20, p = .004, \text{adj. } \eta_p^2 = .15$, indicating that the set size effects (Set Size 8 minus Set Size 2) were larger for relevant cues than irrelevant cues. There was also a significant three-way interaction of set size, cue validity, and cue relevance, $F(1, 47) = 5.12, p = .028, \text{adj. } \eta_p^2 = .01$, indicating the two-way interaction cue validity and cue relevance (i.e., the contingent capture effect) was greater at Set Size 8 than Set Size 2.

Preplanned one-sample t tests assessed the significance of each cue validity effect against zero for each cue type and cue relevance. For irrelevant cues, cue validity effects were significantly greater than zero for both Set Size 2 (12 ms), $t(47) = 3.67, p < .001, d_z = 0.53$, and Set Size 8 (19 ms), $t(47) = 4.21, p < .001, d_z = 0.61$. For relevant cues, cue validity effects were significantly greater than zero for both Set Size 2 (41 ms), $t(47) = 8.83, p < .001, d_z = 1.28$, and Set Size 8 (61 ms), $t(47) = 12.55, p < .001, d_z = 1.81$.

Probability of capture

Estimates of the probability of capture are depicted in Fig. 6. As can be seen, the basic pattern is similar to Experiment 1. A paired-sample t test compared the probabilities of capture

between the two cue types. The probability of capture by relevant cues (36.3%) was significantly greater than that for the irrelevant cues (11.6%), $t(47) = 2.46, p = .017, d_z = 0.36$ (Fig. 6C). Preplanned one-sample t tests were conducted to compare the probability of capture against zero for each cue type (relevant and irrelevant). The probability of capture by relevant cues was significantly greater than zero, $t(47) = 3.80, p < .001, d_z = 0.55$, whereas the probability of capture by irrelevant cues was not, $t(47) = 1.48, p = .15, d_z = 0.21$.

Error rates

The same analyses on manual RT were repeated for error rates, which were quite low (3.5%). There was a main effect of cue validity, $F(1, 47) = 28.00, p < .001, \text{adj. } \eta_p^2 = .36$, indicating error rates were higher for invalid cues than valid cues. There was also a main effect of set size, $F(1, 47) = 7.68, p = .008, \text{adj. } \eta_p^2 = .12$, indicating a larger error rate for Set Size 2 than Set Size 8. The interaction for cue validity and cue relevance was trending significant, $F(1, 47) = 3.35, p = .074, \text{adj. } \eta_p^2 = .05$, indicating a trend of a larger cue validity effect on error rates for relevant cues than irrelevant cues. All other main effects and interactions were nonsignificant ($ps > .10$).

Discussion

Experiment 2 used a more conventional experimental approach in which each participant completed a single session. To compensate for the lower number of trials per participant, we tripled the number of participants and again used the jackknife technique to reduce measurement error. The results replicated the key findings of Experiment 1. There was a smaller cue validity effects for irrelevant cues than relevant cues, as predicted by a contingent capture account (Folk et al., 1992).

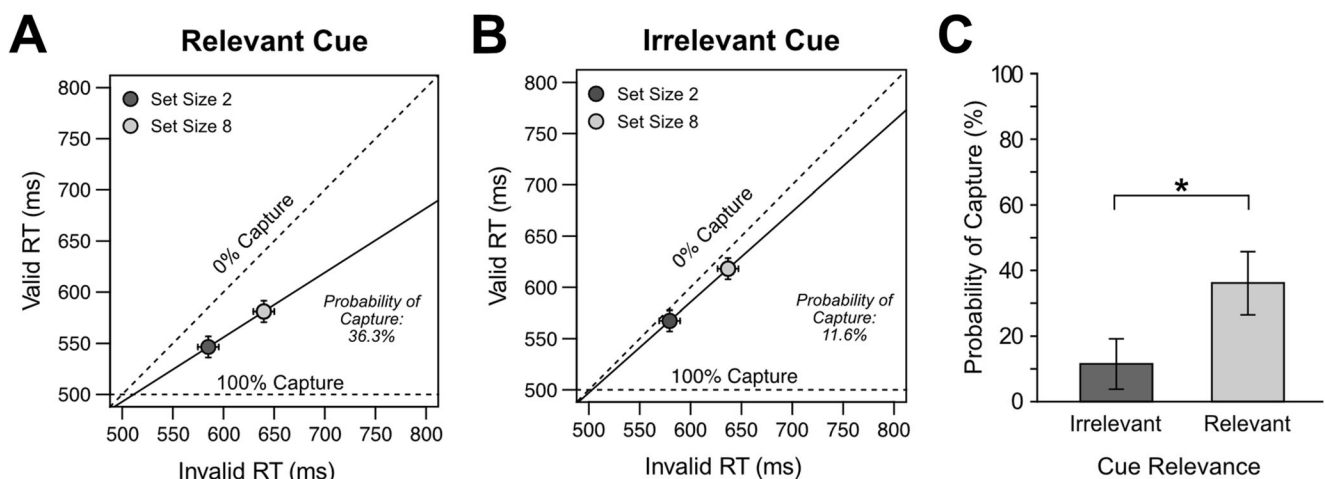


Fig. 6 Results for Experiment 2. Response time on valid versus invalid trials for (A) relevant cues (target color) and (B) irrelevant cues (nontarget color). C A comparison of the estimated probability of capture for

relevant cues and irrelevant cues. Error bars represent 95% within-subject confidence intervals

Importantly, the probability of capture also showed a contingent capture effect with a higher probability of capture for the relevant cues (36.3%) than the irrelevant cues (11.6%). As in Experiment 1, even capture by relevant cues (matching the target color) did not approach 100% as might be expected by some theories of attentional capture.

General discussion

Most metrics used to study attentional capture are latency-based (Folk et al., 1992; Theeuwes, 1992; Yantis & Jonides, 1984). Although these metrics are useful because they can indicate whether capture occurred, they also have an important limitation: they do not directly indicate how frequently capture occurred. The current study sought to address this problem by introducing a novel method of data analysis that can estimate probability of attentional capture. Our probability of capture metric can be illustrated graphically by plotting invalid RT by valid RT for each set size (Fig. 2). The slope of the line on such plots can be used to estimate the underlying probability of attentional capture, indicating how powerful salient and/or relevant stimuli are at attracting attention.

Experiment 1 estimated the probability of capture in a modified spatial cueing paradigm. A multisession approach, in which each participant completed many trials over several sessions, was used to provide a more reliable estimate of the probability of capture. The results demonstrated the typical contingent capture effect of larger cue validity effects for relevant-colored cues than irrelevant-colored cues. Importantly, estimates of the probability of capture were higher for relevant cues (28.5%) than irrelevant cues (4.5%), complementing the standard cue validity effects based on mean RT. This result demonstrates that contingent capture does influence the probability of capture, but that the overall probability of capture was lower than expected.

Experiment 2 was a replication of Experiment 1 with a more traditional and practical sampling approach in which each participant completed a single experimental session. To reduce measurement error in the estimates of probability of capture, the sample size was increased to 48 participants and jackknifing was used. All other methodological details were identical to Experiment 1. Once again, the estimate of probability of capture was higher for relevant cues (36.3%) than irrelevant cues (11.6%), replicating the pattern obtained in Experiment 1.

Two main conclusions can be drawn from the current study. First, the results indicate that the *probability* of attentional capture is strongly influenced by the task-relevance of a distracting stimulus, consistent with a contingent capture account (Folk et al., 1992; Folk & Remington, 1998). Attentional capture was four times more likely to occur from relevant cues (32% capture, averaged across both

experiments) than irrelevant cues (8% capture). Although there was some evidence of capture by irrelevant cues, the probability of capture was very weak. This demonstrates a benefit of the probability of capture metric: the probability of capture estimate allows one to contextualize the relative strength of the capture compared with maximal capture. For example, one might be tempted to conclude from these experiments that irrelevant cues captured attention because they produce a small cue validity effect that is greater than zero. But the probability of capture estimate (8%) shows that capture was negligible.

A second conclusion is that attentional capture is generally less powerful than one might assume. Even capture by relevant cues—which are assumed to be among the most potent of stimuli—was relatively weak (32%, averaged across both experiments). In other words, relevant cues captured attention on only a third of trials. Importantly, a control condition with a predictive cue suggested that our method was sufficiently sensitive to detect 100% capture when it actually occurs (i.e., with informative cues that reliably predict the target location). Thus, involuntary attentional capture is not nearly as powerful as endogenous control of attention by predictive cues (see also Maxwell et al., 2021; Yantis & Jonides, 1990). It is conceivable that something about the current experimental design somehow undermined capture. For example, this particular paradigm may have not induced a strong top-down attentional set, yielding a lower probability of capture than in other paradigms. But we will address this in the next section by demonstrating that the current estimate of probability of capture is quite similar to those derived from previous studies.

Estimating probability of capture in previous studies

Importantly, our new estimate the probability of capture permits a meaningful comparison of the strength of the capture across studies, even if they used different stimuli and different paradigms. To demonstrate this, we estimated the probability of capture from previously published studies. Studies were included if (a) they included a set size manipulation, and (b) they included both valid and invalid cues (nonpredictive of target location). We used either mean RT values reported directly in the paper or, failing that, those estimated from the figures using online tools for extracting values from graphs. The estimates for each of the 27 experiments can be found in Table 2.

Figure 7 depicts the average estimate of probability of capture for each type of salient stimulus (e.g., irrelevant color singletons, relevant color singletons, onsets). As can be seen, the probability of capture estimates for the irrelevant color singletons (14%) and relevant color singletons (42%) were quite close to the values observed in the current study (8% and 32%, respectively). Interestingly, abrupt onsets seem to be generally more potent at attracting attention than the other

Table 2 Probability of capture estimates from previous studies

Article	Exp.	Salient Stimulus	Probability of Capture
Jonides and Yantis (1988)	1	Onset	95%
Jonides and Yantis (1988)	1	Luminance	0%
Jonides and Yantis (1988)	1	Irrelevant Singleton	–38%
Jonides and Yantis (1988)	2	Onset	75%
Jonides and Yantis (1988)	2	Luminance	–20%
Jonides and Yantis (1988)	2	Irrelevant Singleton	8%
Yantis and Egeth (1999)	3	Irrelevant Singleton	27%
Yantis and Egeth (1999)	5	Motion	38%
Yantis and Egeth (1999)	8	Motion	24%
Yantis and Egeth (1999)	8	Irrelevant Singleton	25%
Yantis and Egeth (1999)	8	Luminance	61%
Franconeri and Simons (2003)	1	Onset	46%
Franconeri and Simons (2003)	1	Irrelevant Singleton	28%
Franconeri and Simons (2003)	2	Motion	63%
Franconeri and Simons (2003)	2	Motion	52%
Franconeri and Simons (2003)	2	Motion	59%
Franconeri and Simons (2003)	3	Motion	43%
Franconeri and Simons (2003)	3	Motion	3%
Gaspelin et al. (2012)	1 & 2	Relevant Singleton	20%
Gaspelin et al. (2012)	1 & 2	Irrelevant Singleton	8%
Gaspelin et al. (2012)	3	Onset	32%
Yantis and Jonides (1984)	1	Onset	65%
Turatto et al. (2004)	3	Irrelevant Singleton	38%
Todd & Kramer (1994)	1	Relevant Singleton	22%
Todd & Kramer (1994)	2	Onset	32%
Yeh & Liao (2010)	1	Onset	92%
Yeh & Liao (2010)	2	Relevant Singleton	83%

Probabilities of capture as calculated from previous studies

types of stimuli. For example, abrupt onsets produce much higher estimates (63%) of capture than irrelevant color singletons (14%) and luminance singletons (14%). This suggests that dynamic stimuli are generally more powerful at attracting attention than color singletons and other static stimuli (Adams et al., 2022; Franconeri & Simons, 2003; Jonides & Yantis, 1988; Ruthruff et al., 2020).

We do urge some caution when interpreting these results. For one thing, many of the tasks may have made the “task-irrelevant” stimuli partially task-relevant via design choices (e.g., encouraging singleton detection mode; Bacon & Egeth, 1994). Furthermore, certain types of stimuli (e.g., abrupt onsets) tended to be studied with certain kinds of paradigms (e.g., the irrelevant feature paradigm) that might artificially inflate capture estimates. For example, the irrelevant feature paradigm may give an unfair perceptual advantage to the target on valid trials because the target alone appears where there had been no premask before.

What exactly is the “probability” of capture?

Having demonstrated the broader applicability of our estimation approach, we now return to a discussion of what exactly the estimate reflects. The simplest conceptual framework, adopted throughout much of this paper and implied by many conceptualizations of capture (e.g., see Lamy, 2021; Leonard, 2021), is that attentional capture is a probabilistic event that results in either full capture or no capture on a given trial (a *mixture model*). In this framework, the current method indicates the probability of full capture occurring on a given trial (e.g., capture will occur on 30% of trials; e.g., see Theeuwes, 1990, Exp. 3). A more complicated possibility, however, is that capture on a given trial is graded (Anderson & Folk, 2010). For instance, capture by a valid cue might bias search towards the target on every trial, but without guaranteeing that the target is always searched first. Anderson and Folk (2010) argued against the mixture model and in favor of graded

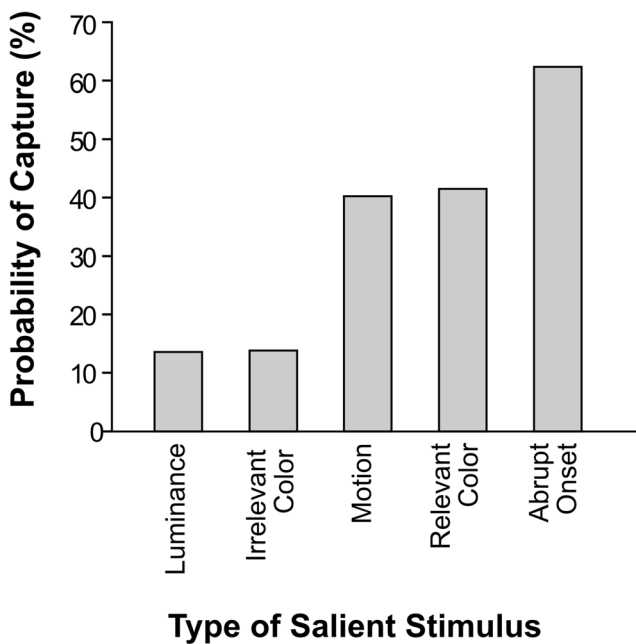


Fig. 7 Probability of capture for different types of salient stimuli estimated from previous studies. These averages were obtained from the studies depicted in Table 2

capture. However, their method required the assumption that task-relevant cues capture attention on 100% of trials. This assumption was plausible at the time, but is directly contradicted by the present data, which show relatively weak capture by task-relevant cues.

These two scenarios (all-or-none capture vs. graded capture) differ meaningfully, but they can be considered functionally equivalent with regard the current metric of capture, which provides an objective estimate of strength of capture, anchored by meaningful endpoints (no capture vs. full capture). In both scenarios, the probability of capture metric would indicate that the relevant cue was strong enough to reduce visual search times by 30%. As an analogy, Cowan's K estimates of working memory capacity is based on the assumption of fixed slots, but has utility as an objective and meaningful measure of capacity even if one relaxes that assumption (Cowan, 2001; Luck & Vogel, 2013).

Relationship to overt attention

Whereas the present methodology provides a way to estimate the probability of capturing covert attention, other methods have been developed for estimating the probability of capturing overt attention. Interestingly, the estimate values are similar. For example, Adams and Gaspelin (2021, Exp. 2) found that 34% of first eye movements were directed to a task-relevant color singleton. This percentage of overt attentional capture is similar to the current estimates of attentional capture by relevant color singletons (32% averaged across both experiments). However, the relationship between covert attention

and eye movements is not entirely understood (Li et al., 2021; Talcott & Gaspelin, 2021). It is possible for covert attention to be captured even when overt eye movements are withheld to prevent oculomotor capture (Kim & Anderson, 2022; van Zoest et al., 2004). Thus, it is not clear that overt attentional capture and covert attentional capture will necessarily yield the same underlying probability of capture.

Considerations for future research

The current metric of probability of capture could lead to new insights about attentional capture. We would, however, like to highlight some important considerations for future researchers using this approach.

1. *Large numbers of trials:* We recommend using many trials per participant, especially if the jackknifing correction is not used. To demonstrate the utility of large numbers of trials, we calculated nonjackknifed probability of capture estimates for relevant cues in Experiment 1 as a function of each added experimental session (Session 1, Sessions 1+2, Sessions 1+2+3). The standard error of relevant cues decreased substantially as each session was added (21.1%, 9.5% and 7.7%, respectively). Thus, more trials will yield more stable estimates of the probability of capture. It is unclear how many trials would be needed, but a single experimental session is likely insufficient for nonjackknifed values.
2. *Jackknifing correction:* We recommend using a statistical technique to appropriately calculate measurement error, such as jackknifing (J. Miller et al., 1998; R. G. Miller, 1974; Oranje, 2006; Ulrich & Miller, 2001) because the probability of capture is a ratio score and can therefore be noisier than other difference score metrics, such as cue validity effects. Jackknifing is particularly beneficial with a limited number of trials. To demonstrate this, we compared standard error of the probability of capture estimate for relevant cues with and without the jackknifing correction in Experiments 1 and 2. In Experiment 1, standard error was roughly equivalent with jackknifing (7.2%) and without jackknifing (7.7%). This is because Experiment 1 had many trials per participant which should minimize measurement error. Experiment 2 had a third as many trials per participant and jackknifing had a much stronger effect. Standard error was much smaller with jackknifing (9.5%) than without jackknifing (32.9%). This demonstrates the utility of the jackknifing correction, especially in situations where measurement error is high due to a low number of trials.
3. *Not ideal for individual differences:* Related to Points 1 and 2, the probability of capture estimate might not be well suited for studies of individual differences, as you cannot reliably measure a person's probability of capture

within a single session. The technique seems to work best on either jackknifed or grand-averaged mean RTs, due to the higher signal-to-noise ratio. Although this might seem like a limitation, we would like to point out that many RT-based metrics of attentional capture have similar issues (Hedge et al., 2018).

4. *Large set size effects*: For the sake of accurately estimating the probability of capture, we suggest using a search task that ensures a large set size effect (i.e., a steep search slope). A large set size effect maximizes the distance between the two points on the invalid-valid RT plots and will therefore permit a more accurate measurement of the slope. Certain capture paradigms typically use visual search tasks that will produce shallow search slopes and these tasks are not ideal for the current estimation technique. Future research is needed to determine whether our approach could work in such paradigms.

Concluding remarks

The present study offers several contributions. The first is a new method of estimating the percentage of trials in which attentional capture occurred. Unlike latency-based metrics of capture, this new metric meaningfully scales the observed strength of capture as a ratio with the maximum potential capture. The second is new data regarding the relative contribution of bottom-up salience and top-down goals in guiding attention. The results indicated that irrelevant color singleton cues produce little-to-no capture, which is consistent with goal-driven theories of attention capture. The third is that relevant cues captured attention on a significantly larger proportion of trials (approximately 30%), but still less than half the time. We found similar values when reanalyzing previous studies to which our method could be applied. At face value, these findings suggest that attentional capture is not as powerful as is typically assumed.

Appendix

Estimating the probability of capture: A mathematical explanation

We denote the baseline visual search time in the absence of spatial cues as ST2 for Set Size 2 and ST8 for Set Size 8. The probability of the spatial precue capturing spatial attention on a given trial, denoted as $p(\text{capture})$, is assumed to be a function of the stimuli, tasks, and participants, but constant across set sizes and cue validity conditions. The most critical assumption is that search time is zero on valid trials in which capture occurs (i.e., capture, by definition, draws spatial

attention straight to the target location, eliminating the need to search any other locations). When capture fails to occur, or occurs but draws attention to a distractor, the search time will be equal to the baseline search time in the absence of a cue (ST2 or ST8). Accordingly, the following equations describe the search time for each set size (2 vs. 8) for each level of cue validity (valid vs. invalid):

$$ST2_{\text{valid}} = p(\text{capture}) * 0 + (1 - p(\text{capture})) * ST2 \quad (4)$$

$$= (1 - p(\text{capture})) * ST2$$

$$ST2_{\text{invalid}} = p(\text{capture}) * ST2 + (1 - p(\text{capture})) * ST2 \quad (5)$$

$$= ST2$$

$$ST8_{\text{valid}} = p(\text{capture}) * 0 + (1 - p(\text{capture})) * ST8 \quad (6)$$

$$= (1 - p(\text{capture})) * ST8$$

$$ST8_{\text{invalid}} = p(\text{capture}) * ST8 + (1 - p(\text{capture})) * ST8 \quad (7)$$

$$= ST8$$

We can then calculate set size effects for valid trials and for invalid trials:

$$ST8_{\text{valid}} - ST2_{\text{valid}} = (1 - p(\text{capture})) * (ST8 - ST2) \quad (8)$$

$$ST8_{\text{invalid}} - ST2_{\text{invalid}} = ST8 - ST2 \quad (9)$$

We can then take the ratio of these sets size effects:

$$\frac{ST8_{\text{valid}} - ST2_{\text{valid}}}{ST8_{\text{invalid}} - ST2_{\text{invalid}}} = 1 - p(\text{capture}) \quad (10)$$

After rearranging, we can solve for the probability of capture.

$$p(\text{capture}) = 1 - \frac{ST8_{\text{valid}} - ST2_{\text{valid}}}{ST8_{\text{invalid}} - ST2_{\text{invalid}}} \quad (11)$$

We further assume that the overall observed RT is equal to the visual search time plus any nonsearch stages of processing (e.g., response selection and execution), which are assumed to not be influenced by validity and set size. Accordingly, probability of capture can be estimated by 1 minus the ratio of the observed set size effects on RT for valid trials and invalid trails.

$$p(\text{capture}) = 1 - \frac{RT8_{\text{valid}} - RT2_{\text{valid}}}{RT8_{\text{invalid}} - RT2_{\text{invalid}}} \quad (12)$$

$$= 1 - \frac{\text{set size effect}_{\text{valid}}}{\text{set size effect}_{\text{invalid}}}$$

Note that this set size effect ratio corresponds to the slope in a plot of valid RT vs. invalid RT (see Fig. 2). Also note that the same math and logic applies to not just to set size effects but would generalize to any manipulation of search difficulty.

This probability of capture equation holds true even if an additional RT cost (k) is incurred following capture on invalid trials, provided that k is independent of set size. For example, it might be especially difficult to disengage from the cued distractor location. Relatedly, consider a serial visual search in which previously searched locations are not revisited. Participants will, on average, search $\frac{1}{2}$ of an item more on invalid trials following capture than on trials without a cue or without capture, regardless of set size. Allowing for the cost k on invalid trials gives the following equations.

$$ST2_{valid} = p(\text{capture}) * 0 + (1 - p(\text{capture})) * ST2 \quad (13)$$

$$= (1 - p(\text{capture})) * ST2$$

$$ST8_{valid} = p(\text{capture}) * 0 + (1 - p(\text{capture})) * ST8 \quad (14)$$

$$= (1 - p(\text{capture})) * ST8$$

$$ST2_{invalid} = p(\text{capture}) * (ST2 + k) \quad (15)$$

$$+ (1 - p(\text{capture})) * ST2$$

$$= ST2 + p(\text{capture}) * k$$

$$ST8_{invalid} = p(\text{capture}) * (ST8 + k) \quad (16)$$

$$+ (1 - p(\text{capture})) * ST8$$

$$= ST8 + p(\text{capture}) * k$$

Importantly, when the set size effects are calculated for invalid trials, the cost k will cancel out and the slope equation will remain the same as above:

$$p(\text{capture}) = 1 - \frac{ST8_{valid} - ST2_{valid}}{ST8_{invalid} - ST2_{invalid}}. \quad (17)$$

Although we see no compelling reason to suspect that a cost of capture on invalid trials would interact substantially with set size, this could be addressed in future studies by replacing the invalid condition with a neutral condition (i.e., cue absent trials).

Funding This project was made possible by National Science Foundation Grant BCS-2045624 to Nicholas Gaspelin. Stimulus programs, data analysis programs, and data are publicly available at <https://osf.io/s94n5/>.

Data availability Stimulus programs, data analysis programs, and data are publicly available on the Open Science Framework (<https://osf.io/s94n5/>).

References

- Adams, O. J., & Gaspelin, N. (2021). Introspective awareness of oculomotor attentional capture. *Journal of Experimental Psychology: Human Perception & Performance*, 47(3), 442–459. <https://doi.org/10.1037/xhp0000898>
- Adams, O. J., Ruthruff, E., & Gaspelin, N. (2022). Oculomotor suppression of abrupt onsets versus color singletons. *Attention, Perception, & Psychophysics*. <https://doi.org/10.3758/s13414-022-02524-0>
- Anderson, B. A., & Folk, C. L. (2010). Variations in the magnitude of attentional capture: Testing a two-process model. *Attention, Perception, & Psychophysics*, 72(2), 342–352. <https://doi.org/10.3758/APP.72.2.342>
- Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics*, 55(5), 485–496.
- Becker, S. I., Folk, C. L., & Remington, R. W. (2010). The role of relational information in contingent capture. *Journal of Experimental Psychology: Human Perception and Performance*, 36(6), 1460–1476. <https://doi.org/10.1037/a0020370>
- Becker, S. I., Folk, C. L., & Remington, R. W. (2013). Attentional capture does not depend on feature similarity, but on target-nontarget relations. *Psychological Science*, 24(5), 634–647.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436. <https://doi.org/10.1163/156856897X00357>
- Chang, S., Niebur, E., & Egeth, H. (2021). Standing out in a small crowd: The role of display size in attracting attention. *Visual Cognition*. <https://doi.org/10.1080/13506285.2021.1918810>
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–114.
- Folk, C. L., & Remington, R. W. (1998). Selectivity in distraction by irrelevant featural singletons: Evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 847–858.
- Folk, C. L., & Remington, R. W. (2008). Bottom-up priming of top-down attentional control settings. *Visual Cognition*, 16(2/3), 215–231. <https://doi.org/10.1080/13506280701458804>
- Folk, C. L., & Remington, R. W. (2015). Unexpected abrupt onsets can override a top-down set for color. *Journal of Experimental Psychology: Human Perception and Performance*, 41(4), 1153–1163.
- 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>
- Folk, C. L., Remington, R. W., & Wright, J. H. (1994). The structure of attentional control: Contingent attentional capture by apparent motion, abrupt onset, and color. *Journal of Experimental Psychology: Human Perception and Performance*, 20(2), 317–329. <https://doi.org/10.1037/0096-1523.20.2.317>
- Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics*, 65(7), 999–1010.
- Franz, V. H. (2007). Ratios: A short guide to confidence limits and proper use. *ArXiv:0710.2024 [Stat]*. <http://arxiv.org/abs/0710.2024>
- Gaspelin, N., Ruthruff, E., Lien, M. C., & Jung, K. (2012). Breaking through the attentional window: Capture by abrupt onsets versus color singletons. *Atten Percept Psychophys*, 74(7), 1461–1474. <https://doi.org/10.3758/s13414-012-0343-7>
- Gaspelin, N., Ruthruff, E., & Lien, M.-C. (2016). The problem of latent attentional capture: Easy visual search conceals capture by task-

- irrelevant abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 42(8), 1104–1120. <https://doi.org/10.1037/xhp0000214>
- Geng, J. J., & DiQuattro, N. E. (2010). Attentional capture by a perceptually salient non-target facilitates target processing through inhibition and rapid rejection. *Journal of Vision*, 10(6), 5. <https://doi.org/10.1167/10.6.5>
- Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 50(3), 1166–1186.
- Hou, X., Harel, J., & Koch, C. (2012). Image signature: Highlighting sparse salient regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(1), 194–201. <https://doi.org/10.1109/TPAMI.2011.146>
- Irons, J. L., Folk, C. L., & Remington, R. W. (2012). All set! Evidence of simultaneous attentional control settings for multiple target colors. *Journal of Experimental Psychology: Human Perception and Performance*, 38(3), 758.
- Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics*, 43(4), 346–354.
- Kievit, R., Frankenhuys, W., Waldorp, L., & Borsboom, D. (2013). Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology*, 4, 513. <https://doi.org/10.3389/fpsyg.2013.00513>
- Kim, A., & Anderson, B. (2022). Systemic effects of selection history on learned ignoring. *Psychonomic Bulletin & Review*, 29(4), 1347–1354. <https://doi.org/10.3758/s13423-021-02050-4>
- Kotseruba, I., Wloka, C., Rasouli, A., & Tsotsos, J. K. (2020). *Do saliency models detect odd-one-out targets? New datasets and evaluations*. ArXiv Preprint ArXiv:2005.06583.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Lamy, D. (2021). The attentional capture debate: The long-lasting consequences of a misnomer. *Visual Cognition*, 29(9), 544–547. <https://doi.org/10.1080/13506285.2021.1904076>
- Lamy, D., Leber, A., & Egeth, H. E. (2004). Effects of task relevance and stimulus-driven salience in feature-search mode. *Journal of Experimental Psychology: Human Perception and Performance*, 30(6), 1019.
- Leonard, C. J. (2021). Consensus emerges and biased competition wins: A commentary on Luck et al. (2020). *Visual Cognition*, 29(9), 560–562. <https://doi.org/10.1080/13506285.2021.1912236>
- Li, H.-H., Hanning, N. M., & Carrasco, M. (2021). To look or not to look: Dissociating presaccadic and covert spatial attention. *Trends in Neurosciences*, 44(8), 669–686. <https://doi.org/10.1016/j.tins.2021.05.002>
- Lien, M.-C., Ruthruff, E., & Johnston, J. C. (2010). Attentional capture with rapidly changing attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, 36(1), 1–16. <https://doi.org/10.1037/a0015875>
- Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, 17(8), 391–400. <https://doi.org/10.1016/j.tics.2013.06.006>
- Luck, S. J., Gaspelin, N., Folk, C. L., Remington, R. W., & Theeuwes, J. (2021). Progress toward resolving the attentional capture debate. *Visual Cognition*, 29(1), 1–21. <https://doi.org/10.1080/13506285.2020.1848949>
- Maxwell, J. W., Gaspelin, N., & Ruthruff, E. (2021). No identification of abrupt onsets that capture attention: Evidence against a unified model of spatial attention. *Psychological Research*, 85(5), 2119–2135.
- Miller, R. G. (1974). The jackknife—a review. *Biometrika*, 61(1), 1–15.
- Miller, J., Patterson, T., & Ulrich, R. (1998). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, 35(1), 99–115.
- Mordkoff, J. T. (2019). A simple method for removing bias from a popular measure of standardized effect size: Adjusted partial eta squared. *Advances in Methods and Practices in Psychological Science*, 2(3), 228–232. <https://doi.org/10.1177/2515245919855053>
- Oranje, A. (2006). Jackknife estimation of sampling variance of ratio estimators in complex samples: Bias and the coefficient of variation. *ETS Research Report Series*, 2006(2), i–23. <https://doi.org/10.1002/j.2333-8504.2006.tb02025.x>
- Palmer, J. (1995). Attention in visual search: Distinguishing four causes of a set-size effect. *Current Directions in Psychological Science*, 4(4), 118–123.
- Remington, R. W., & Folk, C. L. (2001). A dissociation between attention and selection. *Psychological Science*, 12(6), 511–515.
- Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, 26(2), 452–467. <https://doi.org/10.3758/s13423-018-1558-y>
- Ruthruff, E., Faulks, M., Maxwell, J. W., & Gaspelin, N. (2020). Attentional dwelling and capture by color singletons. *Attention, Perception, & Psychophysics*, 82(6), 3048–3064. <https://doi.org/10.3758/s13414-020-02054-7>
- Schönhammer, J. G., Becker, S. I., & Kerzel, D. (2020). Attentional capture by context cues, not inhibition of cue singletons, explains same location costs. *Journal of Experimental Psychology: Human Perception and Performance*, 46(6), 610.
- Stilwell, B. T., & Gaspelin, N. (2021). Attentional suppression of highly salient color singletons. *Journal of Experimental Psychology: Human Perception and Performance*, 47(10), 1313–1328. <https://doi.org/10.1037/xhp0000948>
- Stilwell, B. T., Egeth, H., & Gaspelin, N. (2022). Electrophysiological evidence for the suppression of highly salient distractors. *Journal of Cognitive Neuroscience*, 34(5), 787–805. https://doi.org/10.1162/jocn_a_01827
- Talcott, T. N., & Gaspelin, N. (2021). Eye movements are not mandatorily preceded by the N2pc component. *Psychophysiology*, 58(6), Article e13821.
- Theeuwes, J. (1990). Perceptual selectivity is task dependent: Evidence from selective search. *Acta Psychologica*, 74(1), 81–99.
- Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, 51(6), 599–606. <https://doi.org/10.3758/BF03211656>
- Todd, S., & Kramer, A. F. (1994). Attentional misguidance in visual search. *Perception & Psychophysics*, 56(2), 198–210. <https://doi.org/10.3758/BF03213898>
- Turatto, M., Galfano, G., Gardini, S., & Mascetti, G. G. (2004). Stimulus-Driven Attentional Capture: An Empirical Comparison of Display-Size and Distance Methods. *The Quarterly Journal of Experimental Psychology Section A*, 57(2), 297–324. <https://doi.org/10.1080/02724980343000242>
- Ulrich, R., & Miller, J. (2001). Using the Jackknife-based scoring method for measuring LRP onset effects in factorial designs. *Psychophysiology*, 38, 816–827. <https://doi.org/10.1017/S0048577201000610>

- van Zoest, W., Donk, M., & Theeuwes, J. (2004). The role of stimulus-driven and goal-driven control in saccadic visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, 30(4), 746–759. <https://doi.org/10.1037/0096-1523.30.4.746>
- Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, 15(4), 160–168. <https://doi.org/10.1016/j.tics.2011.02.005>
- Yantis, S., & Egeth, H. E. (1999). On the distinction between visual salience and stimulus-driven attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, 25(3), 661–676. <https://doi.org/10.1037/0096-1523.25.3.661>
- 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.
- Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance*, 16(1), 121–134. <https://doi.org/10.1037/0096-1523.16.1.121>
- Yeh, S. L., & Liao, H. I. (2010). On the generality of the displaywide contingent orienting hypothesis: Can a visual onset capture attention without top-down control settings for displaywide onset? *Acta Psychologica*, 135(2), 159–167. <https://doi.org/10.1016/j.actpsy.2010.05.013>
- Zivony, A., & Lamy, D. (2018). Contingent attentional engagement: stimulus- and goal-driven capture have qualitatively different consequences. *Psychological Science*, 29(12), 1930–1941. <https://doi.org/10.1177/0956797618799302>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.