Salience from the decision perspective:
You know where it is before you know it is there

Michael Zehetleitner
Ludwig-Maximilians-Universität München, Munich, Germany
Hermann J. Müller
Ludwig-Maximilians-Universität München, Munich, Germany, & Birkbeck College, University of London, London, UK

In visual search for feature contrast ("odd-one-out") singletons, identical manipulations of salience, whether by varying target–distractor similarity or dimensional redundancy of target definition, had smaller effects on reaction times (RTs) for binary localization decisions than for yes/no detection decisions. According to formal models of binary decisions, identical differences in drift rates would yield larger RT differences for slow than for fast decisions. From this principle and the present findings, it follows that decisions on the presence of feature contrast singletons are slower than decisions on their location. This is at variance with two classes of standard models of visual search and object recognition that assume a serial cascade of first detection, then localization and identification of a target object, but also inconsistent with models assuming that as soon as a target is detected all its properties, spatial as well as non-spatial (e.g., its category), are available immediately. As an alternative, we propose a model of detection and localization tasks based on random walk processes, which can account for the present findings.

Keywords: salience, detection, localization, decision chains


Introduction

In order to perform goal-directed actions to objects in our visual environment, humans and other animals need to know that a relevant target object is there (detection), where it is (localization, e.g., for a directed eye or manual reaching movement), and frequently also what it is (identification, e.g., for manual grasping movements). However, exactly how the processes of detecting, localizing, and identifying a visual object interrelate is as yet little understood—owing to a number of limitations.

First, most studies in the literature focused on the relationship between identification and either detection or localization (18 out of 23 studies found; e.g., Grill-Spector & Kanwisher, 2005), whereas only a few studies have examined the relationship between detection and localization (five out of 23 studies), and those that did proposed all possible combinations of interdependence (see Table 1).

Second, the way detection was operationalized in some of the few studies of detection and localization was problematic (e.g., Green, 1992; Sagi & Julesz, 1985): Neither Green (1992) nor Sagi and Julesz (1985) operationalized detection in terms of a “yes/no” detection paradigm; that is, their tasks did not involve any target-absent trials—arguably a crucial condition for detection paradigms. Instead, Green (1992), for example, used a two-interval forced-choice paradigm for detection, which is basically a temporal localization task (as is acknowledged in Green’s Footnote 1) and, in terms of signal detection theory, structurally similar to a spatial localization task (Macmillan & Creelman, 2005).

Third, most studies examined performance using brief display durations (21 out of 23 studies), which, by pushing the visual system to its limits, provide useful conditions for psychophysical investigations. However, the visual world is relatively stable, permitting inspection times in excess of, say, the 17–70 ms used by Grill-Spector and Kanwisher (2005), which raises questions about the ecological validity of these studies. Such questions appear justified, as, for example, feature singletons can be localized only roughly with brief display presentations, whereas fine localization requires longer display durations (Solomon & Morgan, 2001). In addition, just-noticeable differences increase and contrast sensitivity decreases with display durations below 1 s (Reitner, Sharpe, & Zrenner, 1992; Schober & Hilz, 1965). Regarding scene perception, task-irrelevant categories may interfere with reporting the “gist” (i.e., the general category; Thorpe, Fize, & Marlot, 1996) of a briefly presented scene—but only with short (50-ms), not with long (200-ms), stimulus presentations (Evans & Wolfe, 2009). Thus, with regard to the relationship between detection, localization, and identification, the...
available findings obtained using brief display durations may not readily generalize to natural viewing conditions. Importantly, this also applies to findings from visual-search studies. For example, Verghese and Nakayama (1994) examined how search performance depended on the number of display items for limited and unlimited viewing conditions. They found that for a horizontal target among vertical distractors, a prototypical “pop-out” target, search performance deteriorated as the number of items increased, but only under limited viewing conditions—under unlimited viewing conditions, search reaction times (RTs) for the same target were independent of set size. Finally, none of the studies listed in Table 1 make statements about the time course of detection, localization, and identification decisions. Only few measured RTs, and those that did acknowledged that RT measures may not permit questions of functional architecture (such as seriality of sub-processes; Saarinen, 1996b; see also Townsend, 1971) to be decided. The main problem using RTs is that such measures are composites of decision and non-decision times and both can be responsible for a difference in the observed speed of task performance. However, every decisional process in the brain is characterized by a time course, including detection, localization, and identification decisions. Given this, the present study was designed to fill this gap in our understanding by investigating the time course of detection and localization decisions.

### Table 1

<table>
<thead>
<tr>
<th>Author</th>
<th>Detection</th>
<th>Localization</th>
<th>Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snyder (1972)</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Treisman and Gelade (1980)</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Nissen (1985)</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Sagi and Julesz (1984)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Sagi and Julesz (1985)</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Atkinson and Braddick (1989)</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Müller and Rabbitt (1989)</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Folk and Egert (1989)</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Johnston and Pashler (1990)</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Green (1992)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Duncan (1993)</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Bennett and Jaye (1995)</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Bloem and van der Heijden (1995)</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Saarinen (1996a)</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Saarinen (1996b)</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Saarinen, Vanni, and Hari (1998)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Baldassi and Burt (2000)</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Donk and Meinecke (2001)</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Nothdurft (2002)</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Cameron, Tai, Eckstein, and Carrasco (2004)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Grill-Spector and Kanwisher (2005)</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Evans and Treisman (2005)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Busey and Palmer (2008)</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 1. List of studies that investigated the relationship between detection, localization, and identification.

### Interrelation of detection and localization

Apart from the methodological difficulties outlined above, the studies that examined the interrelation of detection and localization arrived at variable conclusions. Intuitively, an interesting object in the visual scene may first have to be detected, before it can be localized. In the psychological literature, three types of interrelations among detection, localization, and identification have been proposed: Feature Integration Theory (e.g., Treisman & Gelade, 1980) assumes that (post-selective) localization can take place only once the target has been detected/identified pre-attentively, where detection/identification are assumed to involve the same processing step. Sagi and Julesz (1984) proposed that before an item can be (post-selectively) identified, it has to be (pre-attentively) detected and localized, where detection and localization do not depend on each other: they found detection and localization performance to be equivalent and better than identification performance for short stimulus presentations. Guided Search (GS; e.g., Wolfe, 1994; Wolfe, Cave, & Franzel, 1989) assumes that, at first, focal attention has to be guided to a specific location and then the selected item can be detected, localized, and/or identified—without explicitly stating how these three processes depend on each other. With respect to detection and localization, the possible combinations are “first detection, then localization,” “independent detection and localization,” or “first
localization, then detection” (a combination consistent with GS if the guidance of focal-attentional selection is regarded as a form of “localization”). Saarinen, Vanni, and Hari (1998) reported the same brain areas to be involved in all three tasks, an occipito-temporal and a parietal–temporal source, with the ERP sources exhibiting statistically equivalent latencies and equivalent peak amplitudes—except for the right occipital–temporal source where activity was smallest with passive-viewing, larger with detection, and largest with localization and identification instructions. Assuming that the maximum amplitude is indicative of the processing resources necessary for the different tasks, detection would require fewer resources than localization and identification.

The decision perspective

Besides the general lack of research on the relationship between detection and localization (see above), until now there have been no investigations of the time course and duration of the underlying “decisions.” Generally, RT measures are regarded as composites of decision and non-decision time components (e.g., sensory and motor processing) and thus do not readily lend themselves to conclusions about task interdependence. However, this limitation can be overcome by employing formal RT models that explicitly model decision and non-decision times (e.g., Ratcliff, 1978). On this background, guided by predictions derived from formal decision models (e.g., Ratcliff, 1978; Smith & Ratcliff, 2004), the present study was designed to reexamine the specific relationship between detection and localization, from among the possible intertask relations among detection, localization, and identification. Additionally, based on these formal models, the present study was intended to provide a paradigm for investigating this relationship under exposure time-unlimited viewing conditions, which is readily extendable to examining the interrelations of detection and, respectively, localization to identification tasks. The approach taken in the present study is apt and timely, as it has been only recently demonstrated how well the so-called “ex-Wald distribution,” predicted by a diffusion model, fits the observed distribution of RTs in visual-search tasks (Palmer, Horowitz, Torralba, & Wolfe, in press).

In more detail, the present study examined how feature singletons (“odd-one-out” targets) of varying salience are detected and localized in visual search. Salience (e.g., Itti & Koch, 2001) was manipulated by varying (i) the similarity between the target and the distractors in the search array (e.g., Bruce & Tsotsos, 2009; Duncan & Humphreys, 1989) and (ii) the number of dimensions (for visual dimensions, see Wolfe & Horowitz, 2004) in which the target differed from the distractors (dimensional redundancy of target definition). Salience-based models of search performance (such as salience summation models; e.g., Itti & Koch, 2001) basically quantify the difference in feature values at a given location in the search array relative to its surround: the higher the similarity between target and surrounding distractors, the lower the target’s salience. Besides target–distractor similarity, salience is influenced by dimensional redundancy of target definition. Virtually all salience summation models assume that overall saliency signals (which guide focal-attentional selection) are derived by summing the local feature contrast values across visual dimensions. Consequently, a feature singleton that differs from distractors redundantly in two dimensions (e.g., a red tilted bar among green vertical bars, which differs in both color and orientation from the distractors) is more salient than a singleton defined in one dimension only (e.g., a red vertical or a green tilted bar among green verticals, see, e.g., Burrows & Moore, 2009; Koene & Zhaoping, 2007; Zehetleitner, Krummenacher, & Müller, 2009; Zehetleitner, Müller, & Krummenacher 2008). Note that both factors (target–distractor similarity and redundancy of target definition) can be made to modulate the “strength” of the feature singleton without affecting its ability to guide visual search; thus, in the present study, these factors were manipulated such that they did not compromise “search efficiency.” Search is assumed to be efficient if the slopes of the function relating search RT to the number of display items is below 5 ms/item (Wolfe, 1998); in any case, if one wants to avoid using a simple slope criterion: “the closer the slope is to 0 ms per item, the more likely it is that the target is defined by a guiding feature” (Wolfe & Horowitz, 2004, p. 3).

This manipulation relies on salience having two dissociable functional effects: determining (i) the probability of selecting a location and (ii) the time necessary for the first selection. Varying feature contrast, or target–distractor similarity, is the most direct way to manipulate salience, as salience is basically a measure of how much a location differs featurally from its surround. Regarding the probability of selection, there have been many demonstrations that the more similar the target becomes to the distractors (i.e., in terms of features), the larger the rise in search time when increasing the number of items in the display (e.g., Duncan & Humphreys, 1989; Nagy & Sanchez, 1990; Nagy, Sanchez, & Hughes, 1990; Wolfe & Horowitz, 2004). However, it has also been demonstrated that even at high levels of feature contrast at which the target “pops out,” further modulations of feature contrast can still modulate response times (e.g., Verghese & Nakayama, 1994; Zehetleitner, Krummenacher, Geyer, Hegenloh, & Müller, in press; Zehetleitner, Krummenacher et al., 2009; Zehetleitner, Proulx, & Müller, 2009). That is, for these levels of feature contrast, the slopes of the functions relating search times to the number of display items were effectively zero (i.e., the target “popped out”), but the intercept of the functions was influenced by salience. Therefore, we varied feature contrast at a level high enough for salience to ensure similar, that is, effectively zero slopes, while still allowing for variation in the intercepts of the search RT/display size functions.
In formal decision models, visual salience quality can be considered the underlying stimulus quality for making search task decisions, corresponding to the drift rate in Ratcliff Diffusion Models (RDM; Ratcliff, 1978). In an RDM, for a binary decision (e.g., target present/absent, left/right), evidence is assumed to accumulate over time (see Figure 1); and as soon it exceeds an upper or a lower boundary, a decision is triggered. Accumulation of evidence is subject to Gaussian noise; therefore, decisions are made with varying speed, and wrong decisions can be made if, by chance, evidence accumulates against the general drift direction for an extended period of time. By linking the concepts of salience and RDMs, it follows that the higher the salience of the target, the higher the drift rate—resulting in shorter decision times and fewer errors (e.g., Thornton & Gilden, 2007).

For examining the relationship between detection and localization, the present study exploited a geometrical property of RDMs (depicted in Figure 1, see also Ratcliff, Thapar, & McKoon, 2003). For a given modulation in drift rate, a, the resulting difference in decision times depends on the overall duration of the decision: the longer the decision takes, the larger is the resulting difference in decision times. The time a decision takes can be prolonged either by a more conservative decision criterion, or by a slower accumulation rate, or—as in the present study—by task. Thus, RDMs predict that differences in decision times induced by differences in drift rate depend on the overall duration of the decision process: the longer the decision process for the to-be-performed task takes, the larger the observable difference in RTs between decisions (see Figure 1).

![Figure 1. Mean accumulation of sensory evidence (y-axis) over time (x-axis). Presented are two base drift rates (solid lines), one high, \( d_1 \), and one low, \( d_2 \). Both base drift rates are increased by an additive factor \( a \) (dashed lines). There are two decision criteria (\( a_1 \)—conservative, and \( a_2 \)—liberal), and the respective differences in decision times between base and increased drift rates (\( X_1 \), \( Y_1 \), and \( Y_2 \)) are depicted. Note that the same difference in drift rates, \( a \), can lead to a large difference in decision times \( Y_1 \) (for slower decisions and conservative criterion) or smaller differences (for faster decisions), either due to a more liberal response criterion (\( Y_1 \) vs. \( Y_2 \)) or to an increased base drift rate (\( Y_1 \) vs. \( X_1 \)).](https://jov.arvojournals.org/content/jov/10/14/3535/fig/1)

### Rationale of the present study

This prediction can be used to investigate the duration of detection and localization decisions: a difference in RTs induced by a manipulation of stimulus quality (e.g., target–distractor similarity or redundancy of target definition) should be equivalent for localization and detection tasks if both decisions take equally long. If the RT difference is smaller for one task, A, than for the other, B, it follows that a task A decision can be made faster than a task B decision. If task A involves faster decision times than task B, performance of A cannot causally depend on performance of B. By contrast, if decision times in both tasks are similar, the RT differences between (i) targets of high versus targets of low similarity (i.e., low vs. high feature contrast) relative to the distractors and (ii) targets defined in one versus two dimensions should be equivalent for both tasks. Experiment 1 was designed to examine the effect of feature contrast and Experiment 2 that of dimensional target redundancy on RTs in detection and localization tasks.

Three possible outcomes might be expected: (i) the effect of feature contrast and redundancy on RTs may be equivalent in detection and localization tasks; (ii) the effects may be smaller in detection tasks; or (iii) smaller in localization tasks. If the effects are indeed smaller in one task, A, than in the other, B, it can be concluded that decision times are faster in task A than in task B. Models assuming that detection is a prerequisite for localization (e.g., Evans & Treisman, 2005; Treisman & Gelade, 1980) would predict faster decision times for detection than for localization. Models assuming independence of both tasks (e.g., Green, 1992; Sagi & Julesz, 1985) would predict equivalent decision times. Based on other evidence, namely, higher performance (accuracy) for localization than for detection tasks under the same stimulation conditions (Cameron, Tai, Eckstein, & Carrasco, 2004), one would predict faster localization than detection decisions (though the signal detection models of Cameron et al. make no explicit statements about the time course of such decisions).

Guided Search assumes that, independently of the search task, a display item has first to be attentionally selected and then identified and either accepted as a target, or rejected as a non-target, before a response can be made (Cave & Wolfe, 1990, p. 231). Applied to search tasks requiring detection and, respectively, localization responses, this architecture predicts faster decision times for target detection than for localization: In both tasks, the selected item has to be identified as a target; in the detection task, this identification would be sufficient to trigger a target-present/absent response; by contrast, in a localization task, an additional decision about the target location is necessary, increasing the overall decision time for localization tasks.

In two control experiments, we examined whether differences in response selection processes or in target prevalence play a role in the modulation of salience...
effects by task that were investigated in Experiments 1 and 2 of the present study, where Experiment 1 was designed to examine the effect of feature contrast and Experiment 2 that of dimensional target redundancy on RTs in detection and localization tasks. Additionally, to further substantiate the predictions deriving from the decision perspective, we reanalyzed the effect of decision speed on salience in the data of two experiments reported previously (Zehetleitner, Krummenacher et al., 2009). Finally, we present a computational model that can account for the findings observed in Experiments 1 and 2.

**Materials and methods**

**Participants**

Nine observers, 4 males, all right handed, aged between 21 and 50 (median 25) years participated in Experiment 1. In addition, 12 new observers, 5 males, all right handed, aged between 21 and 49 (median 22.5) years took part in Experiment 2.

**Apparatus**

Stimuli were presented on a Sony Multiscan E250 17” monitor driven by a personal computer (PC) with Windows XP operating system, situated in a sound-isolated room with black interior and a dim background light to prevent reflections on the monitor. The experimental software was purpose-written in C++. Viewing distance was about 60 cm. Observers responded by pressing one of two buttons of a custom-made response box (connected to the PC’s parallel port) using the index finger of their left or right hand, respectively. RTs and response accuracy were recorded online by the PC.

**Stimuli and timing**

The stimulus displays used in the two experiments are illustrated in Figure 2. In both experiments, distractors were 34 gray upright bars presented on a black background (<0.1 cd/m²). The bars were arranged around three (invisible) concentric circles of radius of 4.5°, 8.5°, and 12.5° visual angle, respectively, with a 0.2° white fixation spot in the center. The inner circle consisted of 6, the middle circle of 12, and the outer circle of 16 items. A target could be placed on the 2, 3, or 4 o’clock position (right half of the display) or the 8, 9, or 10 o’clock position (left half) of the middle circle (which contained 12 items in total). The bar stimuli extended 0.3° in width and 1.3° in height. Targets were presented on the middle circle only to (i) keep their eccentricity constant (Carrasco,

![Figure 2](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933538/)
maintaining efficient search even for low feature contrast. A set size experiment confirmed search to be efficient for all four types of targets: the 95% confidence intervals of the slope estimates ranged from 0.1 to 1.2 ms/item (high orientation contrast), −0.1 to 1.1 ms/item (high luminance contrast), 0.6 to 3.0 ms/item (low orientation contrast), and 1.1 to 3.0 ms/item (low luminance contrast). In Experiment 2, targets were defined either by feature contrast in one dimension (luminance or orientation) or by feature contrast in both dimensions (luminance and orientation). The feature values of the targets used in Experiment 2 were identical to the high-contrast targets in Experiment 1.

**Design and procedure**

The experiments consisted of two sessions on two consecutive days, each dedicated to a different task. In one session, observers had to indicate presence versus absence of a feature singleton; in the other session, they indicated whether the feature singleton was presented in the left or the right half of the display. The order of tasks was counterbalanced across observers.

**Experiment 1.** In each session, there were 50 feature singleton targets for each combination of dimension (luminance or orientation) and feature contrast (low or high). In the detection session, 200 additional target-absent trials were randomly interleaved among the target-present trials.

**Experiment 2.** In each session, there were 88 targets defined singly by luminance, singly by orientation, and redundantly by luminance and orientation, each. In the detection session, 264 additional target-absent trials were randomly interleaved into the trial sequence.

**Results**

**Experiment 1**

RTs faster than 150 ms and slower than 1200 ms as well as error response RTs were excluded from analysis (<1.1%). *Mean* RTs were subjected to an ANOVA with the factors task (detection vs. localization), contrast (high vs. low), and dimension (luminance vs. orientation). All three main effects were significant: target localization was faster than target detection (413 vs. 494 ms), \( F(1,8) = 77.9, p < 0.0001 \); RTs were faster for luminance-than for orientation-defined targets (447 vs. 460 ms), \( F(1,8) = 9.1, p < 0.02 \), and faster for high-than for low-feature-contrast targets (436 vs. 471 ms), \( F(1,8) = 23.1, p < 0.001 \). Furthermore, the interaction dimension \( \times \) contrast was significant \( (F(1,8) = 19.3, p < 0.002) \) and, more importantly, the interaction task \( \times \) contrast, \( F(1,8) = 7.5, p < 0.03 \); the feature contrast effect was smaller for target localization (399 vs. 426 ms) than for target detection (472 vs. 515 ms) (see Figure 3).

Error rates were lower for the localization than for the detection task (1.4% vs. 4.1%), \( F(1,8) = 13.0, p < 0.007 \), and lower for high- than for low-contrast targets (1.3% vs. 4.2%), \( F(1,8) = 11.5, p < 0.009 \). The task \( \times \) contrast interaction approached significance, \( F(1,8) = 3.7, p < 0.09 \), with contrast having a smaller effect for the localization (0.4% vs. 1.3%) than for the detection task (1.6% vs. 6.9%). The error pattern thus reinforces the RT results.

In summary, localization responses were faster and more accurate than detection responses. Importantly, the modulations of RTs and (as a tendency) error rates induced by the feature contrast manipulation were smaller for the localization than for the detection task. Thus, decision times were shorter for target localization than for target detection.

**Experiment 2**

After applying the same cut-off criteria as in Experiment 1 (leading to the rejection of <0.8% of all trials), mean RTs were analyzed by a repeated-measures ANOVA with the factors task (detection vs. localization) and target type (luminance, orientation, redundant). There was a significant main effect of task, with localization being faster than detection (407 vs. 471 ms), \( F(1,11) = 19.2, p < 0.001 \), and a significant main effect of target contrast. Furthermore, there was a significant main effect of target type, with luminance targets being processed faster than orientation targets (410 vs. 494 ms), \( F(1,11) = 24.4, p < 0.001 \), and a significant interaction between target type and dimension, \( F(1,11) = 11.5, p < 0.005 \).

**Figure 3.** The size of significant RT modulations induced by salience; bars denote 95% confidence intervals. (a) The present Experiments 1 and 2. (b, c) Experiments 1 and 2 of Zehetleitner, Krummenacher et al. (2009). \( \Delta t_c \) denotes the RT difference between low- and high-feature-contrast targets, RSE denotes the RT difference between targets defined redundantly in two versus singly in one dimension, and DSC denotes the RT difference between target dimension changes and repetitions across trials. Decision time was modulated by task in the present study and by feature contrast and, respectively, speed-accuracy trade-off in Experiments 1 and 2 of Zehetleitner et al.
type, \( F(2,22) = 7.1, p < 0.004 \). Redundant targets were responded to faster (423 ms) than luminance-only and orientation-only targets (446 and 448 ms, respectively). Importantly, the interaction between task and target type was also significant: \( F(2,22) = 7.1, p < 0.004 \). Redundant-signal effects (RSEs) were calculated conservatively by subtracting the RTs for redundant targets from the RTs for the numerically faster non-redundant (either orientation or luminance) targets for each observer and task. The mean RSE across both tasks was 19 ms (10 ms for the localization and 28 ms for the detection task, see Figure 3). For both tasks, the RSE was significantly greater than zero, \( t(11) = 3.9, p < 0.003, \) for detection and \( t(11) = 2.6, p < 0.03, \) for localization. Furthermore, the RSE was significantly greater for the detection than for the localization task, \( t(11) = 6.7, p < 0.025. \)

Error rates were low overall (3.5%) and showed no indication of a speed–accuracy trade-off (in the RT effects): error rates were lower for redundantly defined than for singly defined luminance and orientation targets (2.1% vs. 4.0% and 3.0%, respectively). In addition, fewer errors were made in the localization task than in the detection task (1.7% vs. 4.4%), \( F(1,11) = 14.3, p < 0.003. \)

In summary, as expected, in both tasks, redundant targets were responded to faster than single targets (Koene & Zhaoping, 2007; Zehetleitner, Krummenacher et al., 2009). Importantly in the present context, the RSE was significantly smaller for localization than for detection tasks. Given that the same stimulation was used in both tasks, the increase in drift rates for redundantly, compared to singly, defined targets would have been equivalent in both tasks. Therefore, the differential RSEs between the two takes can be taken to indicate that it took less processing time to reach a localization decision compared to a detection decision.

**Intermediate summary**

Taken together, the results of Experiments 1 and 2 suggest that decision times are faster for target localization than for target detection—at variance with models of visual search that assume localization to (serially) depend on completed detection (Evans & Treisman, 2005), or that assume localization and detection to be independent of each other, but to have similar time courses (Green, 1992; Sagi & Julesz, 1985; to some extent, Wolfe, 1994). However, interpretation of the results of Experiments 1 and 2 is difficult as there may be two (potentially) confounding factors at work. First, in detection tasks, the stimulus-response mapping is neutral in spatial congruency; by contrast, in localization tasks, a stimulus on a given side (e.g., the left) is always associated with a congruent (left) response. Thus, localization RTs might be faster than detection RTs because response selection is easier for localization than for detection.

On the other hand, in the detection task, targets were present on only 50% of trials, as compared to 100% in the localization task. Given that target prevalence affects search performance (Wolfe, Horowitz, & Kenner, 2005), it is possible that differential target prevalence, rather than a difference in decision times, between both tasks is responsible for the observed effects. In order to test both alternative accounts, two additional experiments were performed.

**Slowing of decision times—Not general slowing**

We demonstrated that the size of salience effects was smaller for localization than for detection tasks and proposed that localization tasks involved shorter decision times than detection tasks. However, it is possible that general slowing of the response, rather than slowing of decision times, is the crucial factor for the difference between detection and localization tasks. Therefore, we manipulated the duration of localization responses at a level other than the decision stage, namely, the response selection stage. The stimuli were identical to the localization task of Experiment 1, except that observers had to apply two stimulus-response mappings in the two halves of the experiment. In one half, they responded with the left (right) hand if the target was located in the left (right) half of the screen. In the other half, the response mapping was reversed: observers responded to a target on the left (right) side with a right (left) button press. Again, we measured the difference in RTs between high- and low-feature-contrast targets, for both the standard and the anti-localization response mapping. If general slowing were responsible for the observed size of salience effects in Experiments 1 and 2, the salience effect should be larger for the anti- than for the standard localization task. If only differences in decision times can modulate the size of salience effects, the difference between the two levels of feature contrasts should be equivalent for both response mappings. Thirteen participants (6 males), aged between 20 and 31 (median: 22) years, all right-handed, participated in this experiment.

RT faster than 150 ms and slower than 1500 ms were excluded from the analysis (<0.27%). Mean RTs of correct responses were subjected to a repeated-measures ANOVA with factors response mapping (standard vs. anti), dimension (orientation vs. luminance), and feature contrast (high vs. low). The manipulation of response mapping had the desired effect: responses were overall slower with the anti-mapping than with the standard mapping (379 vs. 412 ms; main effect of task, \( F(1,12) = 7.8, p < 0.02 \)). Furthermore, RTs were slower to low-contrast than to high-contrast targets (386 vs. 405 ms; main effect of contrast, \( F(1,12) = 46, p < 0.0001 \)).
Importantly, however, this difference was of the same size for both tasks: the task \times contrast interaction was far from significant ($F < 0.01, p > 0.9$). Additionally, there was a significant main effect of dimension ($F(1,12) = 6.7, p < 0.02$), which interacted with contrast ($F(1,12) = 28.7, p < 0.001$), due to luminance targets being detected somewhat faster than orientation targets (391 vs. 400 ms), while the effect of feature contrast was greater for orientation than for luminance targets (17 vs. 11 ms).

An analogous ANOVA of error rates revealed the same main effects as for the RT data: $F(1,12) = 12.6, p < 0.004$, $F(1,12) = 10.4, p < 0.007$, $F(1,12) = 5.2, p < 0.04$, and $F(1,12) = 5.9, p < 0.03$, for the main effects task, contrast, and dimension, and the dimension \times contrast interaction, respectively. The interaction between contrast and task was non-significant ($F(1,11) = 0.04, p > 0.66$). Error rates were higher for low- than for high-contrast targets (5.3% vs. 2.6%) and for the anti- than for the standard localization task (4.8% vs. 3.1%). That is, there was no indication of a speed–accuracy trade-off.

This pattern of effects confirms that the differential salience effects between the localization and detection tasks in Experiments 1 and 2 can indeed be attributed to differences in decision times, rather than to processing differences at the stage of response selection.

RTs faster than 150 ms and slower than 1500 ms were excluded from the analysis (<0.7%). Mean RTs of correct responses were subjected to an ANOVA with the factors task (100% vs. 50% target-present localization) and feature contrast (high vs. low). Introducing target-absent trials into a localization task indeed slowed RTs compared to 100% target-present localization (464 vs. 421 ms; main effect of task, $F(1,9) = 9.2, p < 0.01$), and low-contrast targets took longer to be responded to than high-contrast targets (466 vs. 418 ms; main effect of contrast, $F(1,9) = 36.0, p < 0.001$). However, the difference between high- and low-contrast targets was numerically identical for both types of localization task (non-significant task \times contrast interaction, $F(1,9)<0.2, p > 0.9$). An analogous ANOVA of the error rates revealed no significant effects at all (all $F < 2.7, p > 0.13$).

Thus, it confirms that the lower frequency of target in the detection compared to the localization tasks of Experiments 1 and 2 was not responsible for the difference in the magnitude of the salience effects.

### Relevance of decision times: Reanalysis of a previous study

The finding that a one and the same manipulation of salience, by feature contrast or dimensional redundancy, leads to smaller RT effects for localization than for detection tasks can be taken to indicate that localization decisions are processed faster than detection decisions. This conclusion follows from the decision perspective principle, namely, that differences in performance induced by differences in stimulus quality (as, e.g., in salience) depend on the duration of the underlying decisions. The study of Zehetleitner, Krummenacher et al. (2009) provides further evidence to substantiate this conclusion. Zehetleitner et al. examined detection of feature contrast singletons defined either singly in one or another dimension or redundantly in two dimensions. In their Experiment 1, single targets could be of high or low feature contrast and, correspondingly, redundant targets too could be of high (high orientation and high luminance contrast) or low salience (though still efficiently detectable; low orientation and low luminance contrast). On the decision perspective principle, the redundant-signals effect (RSE; a difference in salience) would be expected to be larger for slow rather than for fast decisions. This is exactly what was found: redundancy gains were substantially larger for low- than for high-contrast (redundant) targets. In their Experiment 2, Zehetleitner et al. manipulated decision times more directly by a speed–accuracy trade-off manipulation. Targets occurred either on 75% or on 50% of all trials. With the higher target frequency, RTs were overall faster and error rates were increased, indicating successful induction of a speed–accuracy
trade-off. From the decision perspective, one would expect the redundancy gains to be greater for the slower, as compared to the faster, decisions, which is again what was found (see Figure 3).

Note that in Experiments 1 and 2 of Zehetleitner, Krummenacher et al. (2009), the target-defining dimension (luminance or orientation) could either repeat or change across two successive trials. This makes it possible to test a further prediction deriving from the decision perspective: According to the dimension-weighting account (e.g., Found & Müller, 1996; Töllner, Gramann, Müller, Kiss, & Eimer, 2008), salience is increased for repetitions and reduced for changes in the target-defining dimension. The decision perspective predicts that the size of the dimension repetition (vs. change) effect should be greater for slower decisions, that is, for low rather than high levels of feature contrast (Zehetleitner, Krummenacher et al., 2009, Experiment 1) and greater for emphasis on response accuracy rather than on speed (Zehetleitner, Krummenacher et al., 2009, Experiment 2). Therefore, for both experiments, we analyzed the dimension switch (vs. repetition) effects (i.e., the difference in RTs between trials n with a change vs. a repetition of the target-defining dimension relative to the preceding trial n − 1) for fast and for slow decisions. For these intertrial analyses, we only analyzed trials on which a singly defined target was presented on successive trials, excluding trials with an error on the preceding trial (less than 1.5% of the data). Recall that in Experiment 1 of Zehetleitner, Krummenacher et al. (2009), decision time was manipulated by a feature contrast manipulation (faster decisions for high than for low feature contrast), and in their Experiment 2 by manipulating the frequency of target-present trials (speed–accuracy trade-off: faster decisions for blocks in which target presence was more likely). For Experiment 1 of Zehetleitner, Krummenacher et al., an ANOVA with the within-subjects factors feature contrast (low, high) and intertrial transition (same, different dimension) revealed, besides the feature contrast effect [F(1,14) = 87.8, p < 0.0001], the main effect of intertrial transition to be significant [F(1,14) = 76.1, p < 0.0001]: there was a significant dimension switch cost of 19 ms. Importantly, intertrial transition interacted significantly with feature contrast [F(1,14) = 6.2, p < 0.025]: dimension switch costs were greater for low- than for high-contrast targets, 28 versus 10 ms.

For Experiment 2 of Zehetleitner, Krummenacher et al. (2009), an ANOVA with the between-subjects factors present/absent ratio (1:1, 3:1) and redundant/single ratio (1:1, 1:2) as well as the within-subjects factor intertrial transition (same, different dimension) revealed, besides the target-present/absent ratio effect [F(1,60) = 4.9, p < 0.03, 18 ms faster detection RTs with 75% target presence], a significant main effect of intertrial transition [F(1,60) = 71.75, p < 0.0001] and an interaction of intertrial transition with present/absent ratio [F(1,60) = 6.98, p < 0.01]: dimension switch costs were significantly greater for the 1:1 than for the 3:1 present/absent condition: 15 versus 8 ms. Thus, in both experiments of Zehetleitner, Krummenacher et al. (2009), the dimension switch cost effects are in line with the decision perspective, which predicts that differences in drift rates yield larger performance differences for slow compared to fast decisions.

Models of localization decisions

Which decision model, then, could account for faster localization than detection decisions? Signal detection theory (SDT) predicts that performance is worse in a yes/no detection task than in a two-alternative forced-choice (2AFC) localization task, because in a localization task there is more information available for the decision: for each half of the display, there is evidence for or against target presence, leading to an increase in performance (see Figure 4). We propose a similar model for RTs and accuracy of localization decisions: whereas detection decisions are based on one diffusor providing a decision about target presence versus absence, localization decisions can be modeled by two diffusors, each deciding about target presence/absence for one half of the display. A localization decision is triggered as soon as one of both diffusors terminates. This model mirrors our empirical results in terms of faster decision times and smaller RT differences for modulations of salience in localization than in detection tasks.

Signal detection

In signal detection theory, on each trial, sensory evidence is compared against a criterion to come to a decision about the stimulus. In yes/no detection tasks, sensory evidence is assumed to be normally distributed with mean zero for target-absent and with mean d' for target-present displays. Under the assumption of normality and equal variance for both distributions, the distance between the signal and the noise distribution is a measure of sensitivity and calculated as follows:

\[
d'_{\text{yes/no}} = z(\text{hit}) - z(\text{fa}),
\]

where hit denotes the hit rate and fa denotes the false alarm rate. For two-alternative forced-choice (2AFC) tasks, where on each trial a signal and a noise stimulus are presented separated either temporally or spatially,

\[
d'_{\text{2AFC}} = \sqrt{2}[z(\text{hit}) - z(\text{fa})],
\]

which is less than \(d'_{\text{yes/no}}\). That is, for identical target signals used in a yes/no detection and, respectively, a
2AFC localization task, the distance between the two respective distributions of sensory evidence (noise vs. stimulus presence in detection and stimulus left vs. stimulus right in localization) is smaller for yes/no than for 2AFC tasks (Macmillan & Creelman, 2005). The general reason for this is that in a yes/no task, there is one source of evidence for solving the task: either the target is present or it is absent. That is, sensory evidence is drawn either from the signal or the noise distribution. In 2AFC tasks, however, the target is present in one half of the display and absent in the other half. There are, thus, two sources of sensory evidence on each trial. Mathematically, in the yes/no task, sensory evidence is one-dimensional and either drawn from the noise distribution (centered around zero) or from the signal distribution (centered around \(d'\)). In 2AFC tasks, both the sensory evidence from the left half and that from the right half of the display are two-dimensional: each provide evidence for target presence or absence in the respective half. Thus, on each trial, sensory evidence is drawn from one of two two-dimensional
distributions, which is either centered around zero and \(d'\) (stimulus left) or, respectively, around \(d'\) (stimulus right) and zero (see Figure 4b). These two two-dimensional distributions both have distance \(d'\) from the origin, but distance \(\sqrt{2d'}\) from each other. The ideal criterion in that case is a line through the origin with slope 1. With \(d' = 1\), accuracy in a yes/no task is about 70%, compared to 80% in a 2AFC localization task.

### Random walk models

A similar architecture can be used for modeling RTs in 2AFC localization tasks. A yes/no task can be modeled as a diffusion process where, in case of target presence, evidence accumulates toward a yes-response boundary \(a\) with rate \(v_+\), and, in case of target absence, toward the no-response boundary 0 with rate \(v_-\). A yes response is triggered as soon as sensory evidence is greater than \(a\), and a no response as soon as sensory evidence is less than zero (see Figure 4c).

A localization task can be modeled as one diffusor for the left and a second diffusor for the right display half (see Figure 4d). On each trial, the target is either in the left or the right half, that is, the drift rate of one diffusor is \(v_+\) and that of the other \(v_-\). This redundancy of signal processing can be used by linking both diffusors in a parallel race (logically by an OR operator): a response is triggered as soon as one of both diffusors terminates. As long as the termination times of present and absent decisions overlap (which they mostly do; e.g., Chun & Wolfe, 1996), such localization decisions benefit from statistical facilitation (Raab, 1962) and are faster than yes/no detection decisions.

The detection and localization model presented in Figure 4 was fit to mean correct RTs and accuracy (across all observers) for high- and low-contrast targets (as well as target-absent trials for the detection task), for both tasks. A numerical optimization algorithm (Nelder & Mead, 1965) minimized least squares between empirical and data of Experiment 1. Thus, a localization model of two coupled yes/no diffusors for the left and right halves of the display shows qualitatively the same data patterns as the empirical data of the present Experiments 1 and 2.

### Summary of findings

Taken together, modulations of salience, whether by manipulation of target–distractor similarity or redundancy of target definition, had smaller effects on RTs when the task required a localization decision rather than a detection decision. It could be ruled out that this effect is due to differences in response selection processes or target prevalence between both tasks. From the principle (derived from formal models of binary decisions) that

---

**Table 2.** List of the best fitting parameters for the simulation of the data of Experiment 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boundary separation (a)</td>
<td>0.19</td>
</tr>
<tr>
<td>Drift rate for low-contrast targets, (v_{low})</td>
<td>0.16</td>
</tr>
<tr>
<td>Drift rate for high-contrast targets, (v_{high})</td>
<td>0.20</td>
</tr>
<tr>
<td>Drift rate for no target, (v_{absent})</td>
<td>0.19</td>
</tr>
<tr>
<td>Non-decision time for detection, (T_{er,\det})</td>
<td>0.36</td>
</tr>
<tr>
<td>Non-decision time for localization, (T_{er,\loc})</td>
<td>0.1</td>
</tr>
</tbody>
</table>

---

**General discussion**

---
identical differences in drift rates yield smaller RT differences for fast than for slow decisions, it follows that localization decisions are reached faster than detection decisions for feature contrast (“odd-one-out”) singletons. Importantly, these modulatory effects are not achieved by a general slowing but only by slowing of decisions: while changing the response mapping in an anti-localization task (left key for right targets and vice versa) led to an RT decrease of a similar magnitude to that of a change of task, the RT difference induced by a modulation of feature contrast was statistically equivalent for both types of stimulus-response mapping. Similarly, the differential salience-based RT modulation between localization and detection tasks cannot be attributed to differences in target prevalence (Wolfe et al., 2005) between detection (50% targets) and localization (100% targets). When comparing a 100% with a 50% target-present localization task (where observers had to withhold their response if no target was present), the RT difference induced by feature contrast was equivalent in both conditions.

Furthermore, decision times can be modulated not only by task (e.g., detection vs. localization) but also by speed–accuracy trade-off (faster decisions with emphasis on response speed and slower decisions with emphasis on accuracy). For example, in the study of Zehetleitner, Krummenacher et al. (2009), with emphasis on response accuracy rather than speed, the RSE as well as the effect of a further modulation of salience—cross-trial repetition versus change of the target-defining dimension (e.g., Found & Müller, 1996; Töllner et al., 2008)—was revealed to be larger for low- than for high-feature-contrast targets (Figure 4). Thus, decision times can be investigated using the decision perspective principle in general: the longer a decision takes (whether manipulated in terms of a task or a speed–accuracy trade-off variation), the larger the observable effect of differences in drift rate.

**Scope and generalization**

The present study demonstrates that left/right localization decisions can be reached faster than yes/no detection decisions. Although localization has also previously been operationalized as a 2-alternative choice (e.g., Donk & Meinecke, 2001; Green, 1992; Saarinen et al., 1998; Sagi & Julesz, 1985), localization in natural environments has more alternatives—which, in previous studies, has been approached by using of 4- or 8-alternative localization tasks (e.g., Baldassi & Burr, 2000; Busey & Palmer, 2008; Cameron et al., 2004; Johnston & Pashler, 1990). The present data and computational model pertain, in the first instance, to 2-choice (detection and localization) tasks. This raises the question of whether and how the present findings can be generalized to localization with n alternatives (n > 2). At first glance, the model would predict that all n-choice localization decisions would be faster than detection decisions, although decision times should approach that of detection as n increases to infinity: each localization decision could be triggered if and when sensory evidence in one of n diffusors reaches the “yes” boundary, or the evidence in all n – 1 other diffusors reaches the “no” boundary. Although the latter would happen in a certain proportion of trials, this proportion would be expected to decrease as n increases. However, it is unclear whether the assumption that all drift rates (for target presence and absence, v_+ and v_-) stay the same in both detection and localization models produces a good fit (as it does for n = 2) for n > 2. For detection, the “yes” decision is based on sensory evidence distributed across the whole display; for a 2-alternative localization decision, across half the display; and for an n-alternative localization decision, across only 1/n of the whole display. Given this, drift rates for “yes” responses could increase as the area sampled by one diffusor becomes smaller. Ultimately, it is an empirical question how decision times in an n-alternative localization task, such as saccadic selection or manual reaching, relate to those of 2-alternative localization or detection tasks and what model can account for the data. Additionally, although the present study demonstrates that 2-alternative localization decisions can be formed faster than yes/no detection decisions, it makes no further statements about how detection and identification relate to identification. Based on the existing evidence, there seems to be some other factors, such as coarse localization, or perhaps associated with only coarse localization (Evans & Treisman, 2005; Thorpe et al., 1996); by contrast, detailed categorial identification may require attentional analysis, thus having fine localization as a prerequisite (Donk & Meinecke, 2001; Sagi & Julesz, 1985).

**Consequences for the interrelation between detection and localization**

**Detection precedes localization**

Thus, with regard to the question at issue, the present findings are at variance with models of visual search and object recognition that assume a serial cascade of first detection followed by localization of a target stimulus (Evans & Treisman, 2005; Treisman & Gelade, 1980). In such a serial cascade, decision times for detection would have been expected to be faster than decision times for localization; accordingly, the feature contrast manipulation should have had a smaller effect for detection than for localization tasks—the opposite of what we actually observed. Alternative models of detection and localization in visual search assume either that both are independent of each other, or that detection (serially) follows localization.

Recently, a further aspect of Feature Integration Theory (Treisman & Gelade, 1980) has been “revived”: according to FIT, detection decisions are based on spatially non-specific dimension modules, and only localization and
identification decisions depend on a salience map. That is, for detection decisions, evidence for target presence is pooled over the whole visual scene, within different dimensional modules. For example, an orientation singleton would lead to high activation in an orientation module, and a luminance singleton to activity in a luminance module (in both cases independently of where the singleton is located). Two recent studies reported findings in accordance with this dual-route (non-spatial dimensional modules, spatial saliency map) architecture: Chan and Hayward (2009) and Mortier, van Zoest, Meeter, and Theeuwes (2010). In both, performance was compared between detection (non-spatial) and localization or compound-search (both inherently spatial) tasks. Compound tasks (e.g., Duncan, 1985) are a form of identification task: a target is present on every trial and observers respond to (i.e., identify) a target attribute that is different from (and varies independently of) the target-defining feature (see also Bravo & Nakayama, 1992). Both Chan and Hayward (2009) and Mortier et al. (2010) reported three types of effect to dissociate between spatial and non-spatial tasks: Chan and Hayward found significant dimension switch effects (e.g., Found & Müller, 1996) in detection but not in localization tasks, and interference from additional singletons (Theeuwes, 1992) in localization but not in detection tasks. Furthermore, Mortier et al. reported pre-cueing of the dimension of the upcoming target to be effective only in detection but not in localization tasks. Both studies argued that these findings supported FIT, in that dimension switching or pre-cueing is only effective when the task can be processed via the (non-spatial) dimensional-module route but not when it requires localization decisions, in which case the task is processed via the salience map route. This dual-route notion has been challenged by findings of interference from additional singletons in detection tasks (Zehetleitner, Proulx et al., 2009) and of dimensional cueing effects in compound-search (Müller & Krummenacher, 2006; Töllner, Zehetleitner, Gramann, & Müller, 2010) and pointing tasks (Zehetleitner, Hegenloh, & Müller, in revision). The present findings point to a reason why dimension-based cueing and switch effects might be smaller in localization than in detection tasks: the decision perspective principle predicts that the same salience modulation leads to smaller observable RT effects when the required decision takes less time. Consequently, given that localization decisions are made faster than detection decisions, the same modulation in salience caused by dimension switching or dimensional cueing would lead to smaller RT effects. In localization tasks, these effects can be rendered so small that they become undetectable. This interpretation is supported by the finding that in localization tasks, dimension switching and pre-cueing effects become observable in RT and accuracy measures when the localization is slowed by reducing salience: Zehetleitner et al. (in press) found dimensional effects in localization tasks for (in terms of search time per item, efficiently discerned) low-salience targets, and for high salience targets under time-limited viewing conditions.

**Detection and localization are independent**

Several studies have suggested that detection and localization (e.g., Green, 1992; Sagi & Julesz, 1985) are independent of each other. Independence here means that the outcome of a localization decision is not necessary for making a detection decision and vice versa. By itself, this does not imply any statement about the time course of both decisions: it is possible that both decisions are independent of each other and either have the same time course or different time courses. The parallel diffusion model we proposed would be of the latter type: detection and localization decisions are logically independent of each other but differ in their temporal dynamics. However, such an architecture predicts that under brief viewing conditions, localization performance should be superior to detection performance (see model specification above), which is at variance with the findings of Green (1992) and Sagi and Julesz (1985) but in line with those reported by Cameron et al. (2004). In light of the present model, Green’s findings regarding detection and localization are little surprising: he operationalized detection using a temporal 2-interval forced-choice task and localization with a spatial 2-alternative forced-choice task. In both cases, there are two sources of evidence for the respective decision: from two temporally or from two spatially separated “intervals,” one of which is known to include a target. It would be expected that a yes/no operationalization of the detection task would lead to localization performance being superior to detection performance for the same target displays. Sagi and Julesz (1985) operationalized detection using a subitization task, where observers had to count how many (one to four) feature singletons were present in a given display. Again, at least one target was present on every trial and spatial signals were necessary for solving the task (the different peaks on a salience map have to be dissociated in order to be counted; see Found & Müller, 1996, Experiment 3), making this version of a “detection” task similar to localization tasks.

**Implications for guided search**

Guided Search (Wolfe, 1994) is a two-stage model where salience is computed spatially in parallel, and salience signals guide focal attention to select one item that is subsequently identified as a target or a non-target. According to Cave and Wolfe (1990, p. 231), no response can be made without this target/non-target decision. With respect to localization decisions, however, Guided Search makes no explicit assumptions, leaving three possibilities: (i) either localization decisions are dependent on detection decisions; (ii) both decisions are computed independently and in parallel; or (iii) localization decisions are a prerequisite for detection decisions. The first possibility
(i) would assume that, first, an item is selected based on its salience; second, it is identified as a target/non-target; and finally, the location of the target is decided upon. As for the second possibility (ii), following attentional selection, both a target/non-target and a localization decision for the selected item could be computed in parallel, with the restriction that a localization response can only be given after the selected item has been identified as a target, even if the localization decision would have terminated earlier. Finally, it is possible that a localization decision could be based on salience signals or on the coordinates on which attentional selection is based—but the localization response, as in (ii), could only be triggered after the selected item has been identified as a target. All three possibilities have in common that a localization decision can be effectively triggered only after detection decisions have been terminated. In variants (ii) and (iii), localization decisions could be as fast or slow as detection decisions; in variant (i), they would be always slower. In any case, none of these variants is in accordance with faster localization than detection decisions.

Given that prominent models of visual search, such as Guided Search, cannot account for the data pattern observed in the present study, our proposal is that the salience map has a central role for all types of task and that the use of salience signals is flexible and task-dependent. In particular, for non-spatial detection tasks, it is sufficient to compress activity on the saliency map into one decision variable, whereas for two-choice localization tasks activity is compressed into two decision variables. Consequently, it would be predicted that attentional selection can be modeled by adding possible locations (diffusors) that independently accumulate salience and race for a selection decision.

Conclusion

In summary, we have found that equivalent modulations of target salience have less (differential) impact on left/right localization than on (yes/no) detection responses, where localization RTs are overall faster than detection RTs. To explain this pattern, we propose a decision perspective view based on RDMs (specifically Ratcliff et al., 2003). On this view, given the same target (salience) signals, localization decisions are computed faster than detection decisions, because the former involve two independent diffusors (with the evidence accumulated by each diffusor racing toward a decision criterion) and the latter only one. For models of visual search, this view implies that the salience map plays a central role in all types of task, but that the way activity on the saliency map is compressed into one or multiple decision variables is flexible and task-dependent.

Acknowledgments

This research was supported by the following grants: German Cluster of Excellence Cognition for Technical Systems (EC 142; H. J. Müller and M. Zehetleitner) and the DFG Research Group FOR480 (H. J. Müller). We thank Dragan Rangelov for valuable discussions on previous versions of the manuscript.

Commercial relationships: none.
Corresponding author: Michael Zehetleitner.
Email: mzehetleitner@psy.lmu.de.
Address: LMU Department Psychologie, Leopoldstr. 13, 80802 München, Germany.

Footnote

Note, however, that although the feature singleton targets introduced in the present study permit efficient search according to RT criteria under unlimited viewing conditions, in light of Verghese and Nakayama’s (1994) findings, we cannot make statements about whether or not the same targets would also yield (near-) independence of search accuracy from set size under limited viewing conditions.

References


Grill-Spector, K., & Kanwisher, N. (1. Jan 2005). As soon as you know it is there, you know what it is. Psychological Science, 16, 152–160.


