The time course of color- and luminance-based salience effects

Isabel C. Dombrowe*, Christian N. L. Olivers and Mieke Donk

Department of Cognitive Psychology, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

Salient objects in the visual field attract our attention. Recent work in the orientation domain has shown that the effects of the relative salience of two singleton elements on covert visual attention disappear over time. The present study aims to investigate how salience derived from color and luminance differences affects covert selection. In two experiments, observers indicated the location of a probe which was presented at different stimulus-onset-asynchronies after the presentation of a singleton display containing a homogenous array of oriented lines and two distinct color singletons (Experiment 1) or luminance singletons (Experiment 2). The results show that relative singleton salience from luminance and color differences, just as from orientation differences, affects covert visual attention in a brief time span after stimulus onset. The mere presence of an object, however, can affect covert attention for a longer time span regardless of salience.

Keywords: attention, vision, time course, salience

INTRODUCTION

A distinct object in the visual field tends to attract our attention (Theeuwes, 1992; Nothdurft, 2002; Wolfe and Horowitz, 2004). One possibility to account for this effect is to assume that the visual system contains a salience map (Koch and Ullman, 1985; Itti, 2000; Itti and Koch, 2001). The salience map is a combined topographical representation of the relative conspicuity of locations in the visual field. It is computed from several separate feature conspicuity maps in a parallel fashion (Itti et al., 1998). Attention then visits the locations in the visual field in order of decreasing salience-related activity.

An important question is how salience effects develop over time. That is, how do they emerge, do they change as a function of time and if yes, how? According to the salience map model, the effects of salience indeed change. Within this model, the active locations representing salient objects are eventually inactivated by a mechanism denoted as inhibition of return (IOR, Posner and Cohen, 1984; Posner et al., 1985; Klein, 2000). This is assumed to be necessary to allow shifts of attention. For example, within the Itti et al. (1998) model implementation, the activity at visited locations is set to 0, so that the next object can be selected.

Recent work by Donk and colleagues has confirmed that the effect of salience indeed changes over time, but their finding suggest a different mechanism. In one study by Donk and van Zoest (2008), observers were instructed to make a rapid eye movement to the more salient of two line segments, as defined by the difference in orientation relative to a background of homogeneously oriented line segments. Even though only the relative salience of the two objects was relevant to the task, its effect decreased considerably as saccade latency increased. At short latencies eye movements were more likely to go to the more salient object, whereas long-latency eye movements were equally likely to be directed to the more and the less salient line segment. Both were more likely to be selected than the background. In other words,

an initial salience bias was replaced by a general tendency to select any outstanding object, irrespective of its salience value. Donk and Soesman (2010) further investigated the time course of salience effects. In their study salience was completely irrelevant to the task. Again the stimulus consisted of two orientation-defined elements embedded in a field of homogeneously oriented background lines. Now the task was to respond to a colored probe presented at three different stimulus-onset-asynchronies (SOAs) after the orientation display. The probe could appear at either the more salient location, the less salient location, or a background location. At the shortest SOA (42 ms), Donk and Soesman found a relative RT benefit for the more salient location over the less salient location, but this effect was again transient: at longer SOAs (158 and 483 ms), the difference between the more and the less salient location disappeared, while both locations retained their advantage over the background locations. Thus, relative salience seems to have an initial expediting effect on segmenting objects from their background, but once segmented, the salience difference between objects no longer affects selection (see also Einhäuser et al., 2008; Itti, 2005, for similar suggestions). Furthermore, these representations do not appear to be suppressed toward or beyond baseline levels (as would be indicative of classic IOR).

Instead, Donk and colleagues (van Zoest et al., 2004; van Zoest and Donk, 2005, 2006, 2008; Donk and van Zoest, 2008) proposed on the basis of these and other studies that a salient object differs from a less salient object not because it generates a different amount of activity in the salience map, but because it generates activity at an earlier point in time (see also Thorpe, 2001; VanRullen, 2002, 2003). As a consequence, visual selection is predicted to be only affected by the relative salience of objects as long as there is differential activity in the salience map. After some time has elapsed, both a salient and a less salient object are equally likely to attract attention since both are equally strongly represented in the system.

Salience is thus perceived as an emergent property of the temporal order, rather than the level, in which objects generate activity in the salience map.

However, the conclusions of Donk and Soesman (2010) were based on a limited set of data points. Moreover, the implications of their results are strongly limited by the fact that salience was defined in the orientation domain only. There are several studies suggesting that the time course of salience may depend on the specific visual dimension at which the contrast occurs. For example, a study by Parkhurst et al. (2002) compared the influence of individual feature dimensions on eye movements and found that color and luminance differences had a stronger influence on fixation locations than orientation differences (at least for images that were not dominated by straight lines). Differences among feature dimensions have also been reported by others (e.g., Carmi and Itti, 2006; Engmann et al., 2009). For example, using short video clips, Carmi and Itti (2006) found that both color and intensity contrast contributed more to stimulus-driven overt selection than orientation contrast, but less so than motion contrast. Moreover, an analysis of the time course revealed that the influence of color contrast decreased over the first saccades, while the influence of luminance contrast increased during that time. Taken together, these studies suggest that whether or not salience effects are transient may well depend on the specific dimension tested. It is therefore important to test whether the transience found for orientation generalize to other dimensions. This was done for color (Experiment 1) and luminance (Experiment 2). In both experiments, observers were instructed to detect and manually respond to a probe, which was defined by motion. As in Donk and Soesman (2010), the probe display was preceded by a display consisting of mainly background elements at various SOAs. This display also contained two singleton objects, one more salient than the other (defined by a color or luminance difference to the background). The subsequent probe could appear at either one of these locations, or at a background location. To allow a closer examination of the time course of relative salience effects, we now used six different stimulus durations ranging from 30 to 960 ms.

We predicted that if the transience of relative salience is not specific to orientation, it should generalize to color and luminance. Moreover, the more precise time course allowed us to assess whether the two differential salience representations indeed have an asynchronous onset, but then merge into the same level of activation, as predicted by the timing account explained above.

GENERAL METHOD
PARTICIPANTS
In total 22 students of the Vrije Universiteit Amsterdam participated for either course credit or 7 €. Of these, eight students participated in Experiment 1 and 14 students in Experiment 2. They all reported having normal or corrected-to-normal visual acuity. None of the participants reported to be color blind.

APPARATUS
An HP Compaq d530 CMT Pentium IV computer running E-Prime (Psychology software tools, 2003) generated the stimuli on a color calibrated Iiyama Vision Master Pro SVGA 120-Hz screen and acquired the necessary response data through the standard keyboard. Display resolution was 1024×768 pixels. The experiment was run in a dimly lit room at a viewing distance of about 80 cm.
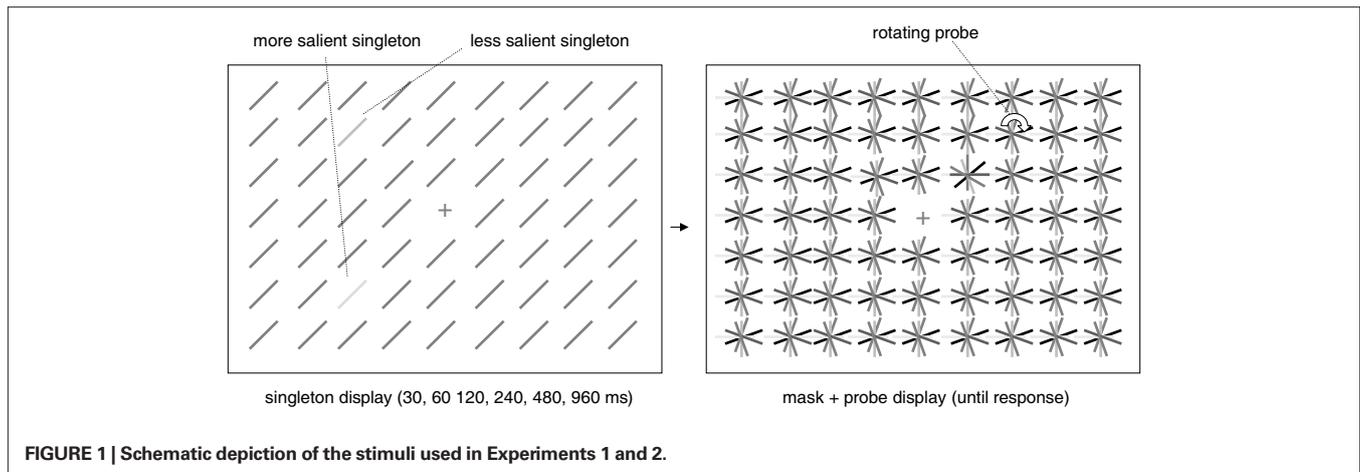
STIMULI
We presented a singleton display followed by a probe display. The singleton display consisted of a field of oriented lines (*background*) and two singleton elements differing from the background elements either in terms of color (Experiment 1) or luminance (Experiment 2). One of the singletons was always more salient (i.e., it was highly distinct from the background elements), the other singleton was always less salient (i.e., it was similar to the background elements). All elements, including the singletons, were 0.8° long, 0.2° wide and always had an orientation of 45° clockwise from the vertical. They were arranged in a 21×29 (*height $\times$ width*) rectangular matrix, with the central element (at position row = 11/column = 15) omitted and replaced by a fixation cross. The singletons were presented on the corners of an imaginary square around the fixation cross (at positions row = 7/column = 11, row = 7/column = 19, row = 14/column = 11, and row = 14/column = 19 of the matrix). These locations all had a distance of 4.3° from the center of the screen. After a variable presentation duration, the singleton display was replaced by a probe display containing multiple asterisk-like elements. Each "asterisk" consisted of six lines of the same length and width with different color (Experiment 1) or luminance values (Experiment 2). One "asterisk" rotated clockwise at about five turns per second (*probe*). The probe was presented at one of the four possible singleton locations. All stimuli were presented on a black background (0 cd/m²). The fixation cross was gray (CIE *x*, *y* coordinates of 0.303; 0.344). **Figure 1** depicts both the singleton display and the probe display.

In Experiment 1, we used two different color singletons, i.e., red (CIE *x*, *y* coordinates of 0.619; 0.342) and green (CIE *x*, *y* coordinates of 0.296; 0.609), and two different background colors, i.e., orange (CIE *x*, *y* coordinates of 0.428; 0.348) and dark green (CIE *x*, *y* coordinates of 0.244; 0.413). We created two singleton displays: a more salient red singleton and a less salient green singleton among dark green background elements and a more salient green singleton and a less salient red singleton among orange background elements. We used the colors of both sets and two additional grays (CIE *x*, *y* coordinates of 0.303; 0.346 and 0.303; 0.344) for the mask elements. All colors were equiluminant (about 12 cd/m²).

In Experiment 2, we used two singletons with different luminance values. The more salient singleton had a luminance of 50.56 cd/m², the less salient singleton a luminance of 30.29 cd/m², and the background elements a luminance of 8.9 cd/m². Thus, the more salient element was the brightest element in the display, whereas the background elements were the least bright. We used three additional luminance values for the mask elements in the probe display; 17.25, 70.67, and 16.98 cd/m².

PROCEDURE AND DESIGN
Each trial started with the presentation of a fixation cross. After 1 s the singleton display appeared on the screen for either 30, 60, 120, 240, 480, or 960 ms (*presentation duration*). Participants were instructed to ignore this display and keep looking at the fixation cross. The display was then masked by a probe display and a short

FIGURE 1 | Schematic depiction of the stimuli used in Experiments 1 and 2.

simultaneous tone (100 ms) was presented, indicating that observers were to report the location of the probe. The singleton locations were not predictive with respect to the probe locations, i.e., on half of the trials the probe was presented at a singleton location (with a quarter of trials on the more salient, the other quarter on the less salient), on the other half it was presented at a background location. Participants were asked to indicate the position of the probe by pressing the "a," "z," "k," or "m" key for "up left," "down left," "up right," or "down right," respectively. A new trial started immediately after the participant's response. Participants were instructed to remain fixated throughout the whole experiment.

Participants first completed 48 practice trials. Both main experiments consisted of 12 blocks of 48 trials, resulting in 576 trials in total. There were four possible probe locations, two singleton locations and two background locations. Thus, there were 144 trials (24 trials per cell) for each singleton location and 288 trials (48 trials per cell) for the background locations. Participants were asked to respond as fast and as accurately as possible. We used a within-subjects design for each experiment.

EXPERIMENT 1

The aim of Experiment 1 was to investigate the temporal dynamics of salience derived from color differences using a probe-RT task. We created two different sets of colors. One of the singletons was always red and one was always green. The color of the background lines differed such that the green singleton was more salient against an orange background and the red singleton was more salient against a green background. Presentation duration was varied across a broad range of values ranging from 30 to 960 ms. We used a motion-defined probe in order to preclude any effects of a top-down attentional set for color (e.g., Folk et al., 1992). Observers had to locate the rotating mask element.

RESULTS AND DISCUSSION

We restricted our RT analysis to trials on which participants had correctly indicated the position of the rotating probe. Reaction times of 2.5 standard deviations above or below the arithmetic mean were excluded for each participant. This resulted in the loss of 5.2% of all trials of Experiment 1. **Figure 2** shows the mean of the remaining RTs for each of the probe locations (more salient, less
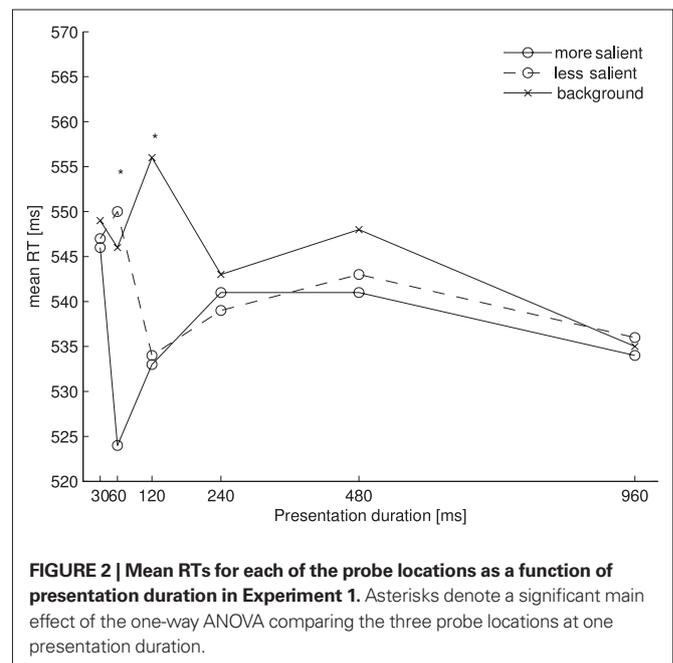
FIGURE 2 | Mean RTs for each of the probe locations as a function of presentation duration in Experiment 1. Asterisks denote a significant main effect of the one-way ANOVA comparing the three probe locations at one presentation duration.

salient, and background) as a function of presentation duration (30, 60, 120, 240, 480, and 960 ms). An overall ANOVA with the same factors revealed a main effect of probe location [$F_{(2, 14)} = 19.567$, $p < 0.001$] and an interaction of probe location with presentation duration [$F_{(10, 70)} = 5.346$, $p < 0.01$]. Participants reacted overall faster to probes presented at more salient locations ($M = 537$ ms) than to those at less salient locations ($M = 541$ ms) or at background locations ($M = 546$ ms). To further assess the interaction we performed separate one-way ANOVAs for each presentation duration. This resulted in significant effects of salience for the durations of 60 ms [$F_{(2, 14)} = 12.968$, $p = 0.001$] and 120 ms [$F_{(2, 14)} = 14.958$, $p < 0.001$], and a marginally significant effect of salience for the duration of 480 ms [$F_{(2, 14)} = 2.985$, $p = 0.083$]. For each duration (including 480 ms) we then performed all three *post hoc* pairwise comparisons, which were Bonferroni-corrected. This revealed significant differences between the more salient and the less salient locations at a presentation duration of 60 ms ($t_{(7)} = 4.316$, $p < 0.01$),

the more salient and the background locations at presentation durations of 60, 120, and 480 ms (60 ms: $t(7) = 4.220$, $p < 0.01$; 120 ms: $t(7) = 3.936$, $p < 0.01$; 480 ms: $t(7) = 4.365$, $p < 0.01$), and the less salient and the background locations at a presentation duration of 120 ms ($t(7) = 5.257$, $p < 0.01$).

Participants made relatively few errors at indicating the location of the rotating probe ($M = 5.4\%$). A within-subjects ANOVA with probe location (more salient, less salient, background) and presentation duration (30, 60, 120, 240, 480, and 960 ms) did not reveal any effects.

We additionally checked if the results for the two displays differed from each other with an ANOVA with display as additional factor. There was neither a main effect of display [$F(1,7) = 2.161$, $p > 0.1$], nor did any interactions with this factor reach significance.

Our results show that the time course of salience from color overall resembles the time course of salience from orientation as described by Donk and Soesman (2010). For both orientation and color, salience effects occurred earlier in time at the more salient location than at the less salient location. Then, salience differences between both singleton locations disappeared while both singleton locations still differed from background locations.

EXPERIMENT 2

In Experiment 2 we investigated how salience defined by luminance contrast dynamically affects visual selection. To this end, we presented observers with displays consisting of a homogenous field of gray oriented lines and two distinct luminance singletons.

RESULTS AND DISCUSSION

Figure 3 shows the mean RTs for each probe location as a function of presentation duration.

The data were analyzed in the same way as in Experiment 1. Removing RTs beyond 2.5 standard deviations from the mean resulted in a loss of 2.7% of all trials. The ANOVA of the individual mean RTs with probe location (more salient, less salient, background) and presentation duration (30, 60, 120, 240, 480,
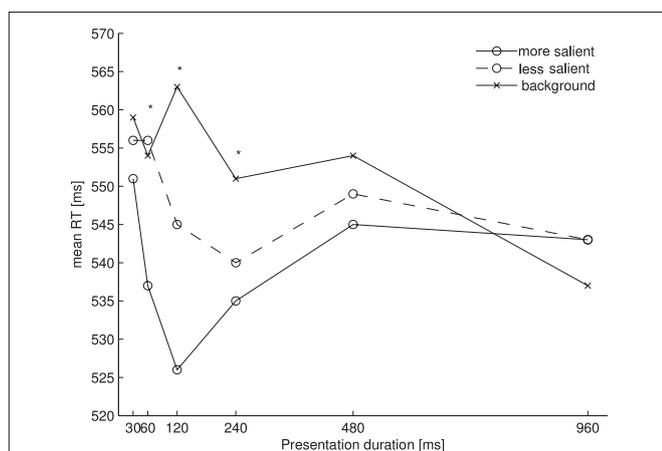

FIGURE 3 | Mean RTs for each of the probe locations as a function of presentation duration in Experiment 2. Asterisks denote a significant main effect of the one-way ANOVA comparing the three probe locations at one presentation location.

and 960 ms) as factors revealed a main effect of probe location [$F(2, 26) = 12.864$, $p < 0.001$] and an interaction of probe location and presentation duration [$F(10, 130) = 3.940$, $p < 0.01$]. Participants reacted faster to probes presented at more salient locations ($M = 539$ ms) than to those presented at less salient locations ($M = 548$ ms) or background locations ($M = 553$ ms). Separate one-way ANOVAs for each presentation duration revealed significant effects of salience for the durations of 60 ms [$F(2, 26) = 4.233$, $p < 0.05$], 120 ms [$F(2, 26) = 22.169$, $p < 0.001$] and 240 ms [$F(2, 26) = 4.923$, $p < 0.05$]. The pairwise comparisons revealed significant differences between the more salient locations and the less salient locations at a presentation duration of 120 ms ($t(13) = 3.486$, $p < 0.01$), between the less salient locations and the background locations at presentation durations of 120 and 240 ms (120 ms: $t(13) = 3.444$, $p < 0.01$; 240 ms: $t(13) = 3.139$, $p < 0.01$) and between the more salient locations and the background locations at presentation durations of 60 and 120 ms (60 ms: $t(13) = 2.901$, $p = 0.014$; 120 ms: $t(13) = 6.199$, $p < 0.001$). At a presentation duration of 60 ms, the comparison between the more salient locations and the less salient locations and the comparison between the more salient locations and the background locations at a presentation duration of 240 ms just failed to reach significance under the Bonferroni corrections (60 ms, more salient vs. less salient: $t(13) = 2.543$, $p = 0.025$; 240 ms, more salient vs. background: $t(13) = 2.653$, $p = 0.02$).

Participants only made few errors in indicating the position of the rotating probe ($M = 4.1\%$). An ANOVA with probe location and presentation duration as factors revealed a main effect of probe location [$F(2, 26) = 6.666$, $p < 0.05$]. Observers made fewer errors when probes were presented at more salient locations ($M = 3.0\%$) than at other locations. Error percentages were similar for less salient and background locations (less salient: $M = 5.0\%$, background: $M = 4.1\%$).

Our results show that the time course of visual salience from luminance contrast is very similar to that from orientation and color contrast. Since we used six presentation durations between 30 and 960 ms we were able to describe the time course of visual salience in more detail than Donk and Soesman (2010). Apparently, salience from luminance contrast only emerges after 30 ms. Then, salience representations are graded, in that a more salient location can be differentiated from a less salient location in the salience map. However, this difference soon vanishes (at presentation durations longer than 120 ms), such that both singleton locations are represented as equally salient. About 480 ms after stimulus onset any salience representations regarding the four probe locations seemed just to have disappeared.

DISCUSSION

The present study aimed to examine the time course of relative color- and luminance-based salience effects. We distinguish relative salience, the difference in salience between two locations that stand out from their surroundings from the mere *presence* of salience, i.e., the fact that these locations stand out from their surroundings. Whether luminance- or color-defined, and whether more or less salient, the presence of a salient object transiently improved probe detection, with performance peaking around 60 ms to 120 ms, after which RTs slowed again. Overall, this time course resembled the time course of spatial cueing effects as reported by Müller and Rabbitt (1989)

and Nakayama and Mackeben (1989). Measuring accuracy, they found optimal performance for a latency of around 100 ms between a salient onset cue and a visual search target, after which accuracy declined again. They suggested that the cue attracted attention, but only transiently. Note that in the current experiments too, overall performance declined again, with RTs at the longest SOA (960 ms) being no different from the background baseline. This overall transience may indeed reflect a passive decay of attention. Alternatively, the disappearance of overall singleton presence effects could be due to top-down influences, possibly reflecting the participants knowledge that singleton locations were uninformative about the location of the upcoming probe. This presentation duration might have been long enough for the participants to ignore the singleton display and indicate the location of the probe dot irrespective of the preceding singleton locations (including background locations).

The most important result was that for both color and luminance, probe RTs were differentially affected by relative singleton salience only briefly after the presentation of the singleton display. Probe RTs to the more salient location were faster than probe RTs to the less salient location or background locations. This RT benefit appeared earlier at the salient location than at the less salient location. However, the different salience effects at the more salient location and at the less salient location were only transiently present: the difference between RTs to probes presented at more salient and less salient locations disappeared after a presentation duration of 60 ms for color and after a presentation duration of 120 ms for luminance. RTs to singleton locations remained faster than to background locations beyond these 60–120 ms, up to about 500 ms when it had disappeared, indicating more sustained effects of object presence than of relative object salience (in line with Itti, 2005; Einhäuser et al., 2008). Eventually, at the longest presentation duration, probe RT was unaffected by either singleton salience or singleton presence.

The time courses of color- and luminance-based salience effects were very similar to the ones found by Donk and Soesman (2010) in the orientation domain. This suggests that the time course of salience effects does not depend on the specific feature dimension. Our results differ from those of the eye movement studies of Parkhurst et al. (2002) or Carmi and Itti (2006), for example, in that they did find feature-based differences. However, one has to keep in mind that these studies investigated the allocation of *overt* attention over the course of several eye movements. Thus, it is possible that different feature dimensions only differentially influence overt selection over the course of several saccades, but not covert selection up to the first saccade.

The sustained benefits for more or less salient objects relative to the background is difficult to explain by a classic IOR mechanism (Posner and Cohen, 1984). Furthermore, an active bias away from the most salient item would predict performance for this location to deteriorate back toward background performance, before the next salient item could be prioritized. Neither appeared to be the case, except the return to baseline levels at the longest presentation duration, but which occurred for both items.

The results obtained in the present study can be more directly explained with a dynamic salience account as proposed by Donk and Soesman (see also Donk and van Zoest, 2008) stating that a more salient location relative to a less salient one does not elicit more activation in the salience map (which would then need to be suppressed), but merely elicits activation at an earlier point in time (see e.g., Thorpe, 2001; VanRullen, 2003, for related proposals). Accordingly, salience can be regarded as an emergent property of the time required for individual locations to maximally activate the corresponding locations in the salience map. This means that the most salient location would be represented first and solely (the more salient location in our experiments). After some time, the second-to-most salient location (the less salient location in our experiment) would also be represented, albeit initially with a lower activation. With a longer stimulus duration, as both items become segmented from the background, activity becomes more and more similar until they are equally active. This way, the time course of visual salience emerges naturally from object segmentation mechanisms.

REFERENCES

Carmi, R., and Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Res.* 46, 4333–4345.

Donk, M., and Soesman, L. (2010). Salience is only briefly represented: evidence from probe-detection performance. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 286–302.

Donk, M., and van Zoest, W. (2008). Effects of salience are short-lived. *Psychol. Sci.* 19, 733–739.

Einhäuser, W., Spain, M., and Perona, P. (2008). Objects predict fixations better than early saliency. *J. Vis.* 8, 1–26.

Engmann, S., 't Hart, B. M., Sieren, T., Onat, S., König, P., and Einhäuser, W. (2009). Saliency on a natural scene background: effects of color and luminance contrast add linearly. *Atten. Percept. Psychophys.* 71, 1337–1352.

Folk, C. L., Remington, R. W., and Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *J. Exp. Psychol. Hum. Percept. Perform.* 18, 1030–1044.

Itti, L. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* 40, 1489–1506.

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Vis. Cogn.* 12, 1093–1123.

Itti, L., and Koch, C. (2001). Computational modelling of visual attention. *Nat. Rev. Neurosci.* 2, 194–203.

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 1254–1259.

Klein, R. (2000). Inhibition of return. *Trends Cogn. Sci.* 4, 138–147.

Koch, C., and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.* 4, 219–227.

Müller, H. J., and Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. *J. Exp. Psychol. Human Percept. Perform.* 15, 315–330.

Nakayama, K., and Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Res.* 29, 1631–1647.

Nothdurft, H.-C. (2002). Attention shifts to salient targets. *Vision Res.* 42, 1287–1306.

Parkhurst, D., Law, K., and Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Res.* 42, 107–123.

Posner, M. I., and Cohen, Y. (1984). Components of visual orienting. *Attent. Perform.* 10, 531–556.

Posner, M. I., Rafal, R. D., Choate, L. S., and Vaughan, J. (1985). Inhibition of return – neural basis and function. *Cogn. Neuropsychol.* 2, 211–218.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Percept. Psychophys.* 51, 599–606.

Thorpe, S. (2001). Spike-based strategies for rapid processing. *Neural Networks*, 14, 715–725.

VanRullen, R. (2002). Surfing a spike wave down the ventral stream. *Vision Res.* 42, 2593–2615.

VanRullen, R. (2003). Visual saliency and spike timing in the ventral visual pathway. *J. Physiol. Paris* 97, 365–377.

van Zoest, W., and Donk, M. (2005). The effects of salience on saccadic target selection. *Vis. Cogn.* 12, 353–375.

van Zoest, W., and Donk, M. (2006). Saccadic target selection as a function of time. *Spat. Vis.* 19, 61–76.

van Zoest, W., and Donk, M. (2008). Goal-driven modulation as a function of time in saccadic target selection. *Q. J. Exp. Psychol.* 61, 1553–1572.

van Zoest, W. V., Donk, M., and Theeuwes, J. (2004). The role of stimulus-driven and goal-driven control in saccadic visual selection. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 746–759.

Wolfe, J. M., and Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nat. Rev. Neurosci.* 5, 495–501.