CORTEX 151 (2022) 89–104

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research Report

Behavioral and ERP evidence that object-based attention utilizes fine-grained spatial mechanisms

Esra N. Catak [a,b], Mert Özkan [c], Hulusi Kafaligonul [a,b,*,1] and Gene R. Stoner [d,1]

[a] Interdisciplinary Neuroscience Program, Aysel Sabuncu Brain Research Center, Bilkent University, Ankara, Turkey
[b] National Magnetic Resonance Research Center (UMRAM), Bilkent University, Ankara, Turkey
[c] Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA
[d] The Salk Institute for Biological Studies, La Jolla, CA, USA

ARTICLE INFO

Article history:
Received 29 June 2021
Reviewed 20 September 2021
Revised 17 December 2021
Accepted 16 February 2022
Action editor Veena Kumari
Published online 15 March 2022

Keywords:
Object-based attention
Transparent-motion
Feature swaps
Color
EEG

ABSTRACT

Valdes-Sosa, Cobo, and Pinilla (1998) introduced a transparent-motion design that provided evidence of *object-based attention* whereby attention embraces all features of an attentionally cued perceptual object including new unpredictable features such as a brief translation. Subsequent studies using variants of that design appeared to provide further behavioral, electrophysiological, and brain imaging evidence of object-based attention. Stoner and Blanc (2010) observed, however, that these previous results could potentially be explained by feature-based competition/normalization models of attention. To distinguish between the object-based and feature-based accounts, they introduced "feature swaps" into a *delayed-onset* variant of the transparent-motion design (Reynolds, Alborzian, & Stoner, 2003). Whereas the object-based attention account predicted that the effect of cueing would survive these feature swaps, the motion-competition account predicted that the effect of cueing would be reversed by these feature swaps. The behavioral results of Stoner and Blanc (2010) supported the object-based account, and in doing so, provided evidence that the attentional advantage in this design is spatially selective at the scale of the intermixed texture elements (i.e., dots) of the overlapping and moving dot fields. In the present study, we used the design of Stoner and Blanc (2010) to investigate both psychophysical performance and evoked activities under different cueing and feature swapping conditions. We confirmed that the behavioral effects of attentional cueing survived feature swaps and found event-related potential (ERP) correlates of those effects in the N1 component range over occipital and parieto-occipital scalp sites. These modulations of the neural activity were, moreover, significantly associated with variation in behavioral performance values across the different conditions. Our findings thus provide the first

* *Corresponding author*. Aysel Sabuncu Brain Research Center Bilkent University, Ankara 06800, Turkey.
E-mail address: hulusi@bilkent.edu.tr (H. Kafaligonul).
[1] Joint senior authors.

https://doi.org/10.1016/j.cortex.2022.02.013
0010-9452/© 2022 Elsevier Ltd. All rights reserved.

evidence of the role of the N1 component in object-based attention in this transparent-motion design under conditions that rule out feature-based mechanisms and that reveal selective processing at a fine spatial scale.

© 2022 Elsevier Ltd. All rights reserved.

1. Introduction

Visual processing is inevitably selective due to the vast amount of information embedded in natural scenes and the limited capacity of the visual system. Selective attention refers to the directing of visual processing to some fraction of the visual input. Feature- and spatially-selective processing is well established and generally consistent with the known selectivity of neurons within the visual pathway. Indeed, normalization and biased-competition models built on these selectivities can parsimoniously account for the neuronal response properties observed in a variety of selective feature-based and spatial attention experiments (e.g., Lee & Maunsell, 2009; Reynolds & Heeger, 2009). Object-based selective processing, conversely, is based on the observation that when one feature of an object is selectively attended, the processing of other features of the same object can also benefit (e.g., Blaser, Pylyshyn, & Holcombe, 2000; Duncan, 1984; Egly, Driver, & Rafal, 1994; Khoe, Mitchell, Reynolds, & Hillyard, 2005; Lopez, Rodriguez, & Valdes-Sosa, 2004; Mitchell, Stoner, Fallah, & Reynolds, 2003; O'Craven, Downing, & Kanwisher, 1999; Pinilla, Cobo, Torres, & Valdes-Sosa, 2001; Reynolds et al., 2003; Rodriguez, Valdes-Sosa, & Freiwald, 2002; Schoenfeld et al., 2003; Valdes-Sosa, Bobes, Rodriguez, & Pinilla, 1998; Valdes-Sosa, Cobo, & Pinilla, 2000; Valdes-Sosa, Cobo, & Pinilla, 1998; Wanning, Rodriguez, & Freiwald, 2007). Unlike feature and spatial attention, the neuronal mechanisms underlying object-based attention are not readily explained by current models.

1.1. *Transparent motion and object-based attention*

Valdes-Sosa, Cobo, and Pinilla (1998) introduced a transparent-motion design as a tool to examine object-based attention absent the potential influence of feature-based or spatial attention. There have been numerous follow-up studies that have used variants of that design to investigate behavioral and/or neuronal correlates of object-based attention. Although there have been some design differences (which we touch upon in *section 4*), all of these studies used stimuli composed of two superimposed counter-rotating, differently colored dot fields. One of the two dot fields is "cued", either endogenously (e.g., fixation point color indicating the color of field to be attended) or exogenously (e.g., by a delayed-onset of one dot field, see *below*). The rotations of the dot fields are interrupted by brief translations (one or two translations, depending upon the design), and subjects are asked to report the direction of those translations. These translations consist of a subset of the dots moving coherently in typically one of eight directions. To discourage tracking of individual dots, the remaining dots of the translating dot field translate in randomly chosen directions. Numerous studies using this basic design have repeatedly found that subjects judge translations of the cued dot field more accurately than translations of the uncued dot field (Khoe et al., 2005; Lopez et al., 2004; Mitchell et al., 2003; Mitchell, Stoner, & Reynolds, 2004; Pinilla et al., 2001; Reynolds et al., 2003; Rodriguez et al., 2002; Stoner & Blanc, 2010; Valdes-Sosa, Cobo, & Pinilla, 1998). In these experiments, spatial attention is ruled out by the spatial intermixing of the moving dots and feature-based (i.e., based on motion direction) selection is ruled out since the direction of the translation is unpredictable. Finally, by removing the color differences between the two dot fields, Mitchell et al. (2003) revealed that the performance bias is not color-based. These findings have thus been taken as evidence of object-based (sometimes referred to as surface-based) attention whereby attention is cued to one of the rotating dot fields, and an attentional benefit extends to a new unpredictable feature (i.e., the translation) of that object.

1.2. *The delayed-onset design*

We can illustrate this idea of object-based attention more concretely in the delayed-onset design (Fig. 1) used in the current study. While simpler than the original design of Valdes-Sosa and colleagues (i.e., fewer translations and responses per trial, and no endogenous attentional cue), Reynolds et al. (2003) provided evidence that this delayed-onset design captures the essence of the original design. In the delayed-onset design, subjects obtain stable fixation followed by the appearance of the first rotating dot field. After a delay, a second (delayed) dot field appears that is rotating in the opposite direction. Following a period of dual rotation, one of the two dot fields (either the first or the delayed dot field) translates briefly (as described *above*), and then both dot fields resume rotation for a brief period of time. Subjects report the direction of the translation at the end of the trial. Reynolds et al. (2003) found that the translations of the delayed dot field were judged more accurately than translations of the dot field that appeared first. As stimulus onsets appear to act as exogenous cues that briefly attract attention (Yantis & Jonides, 1984; 1990), Reynolds et al. (2003) postulated that the onset of the delayed dot field acted as an exogenous cue that captures attention[2] and yields an object-based processing benefit if the delayed dot field translates a few hundred milliseconds later. Their results were interpreted as providing further evidence of object-based selection in transparent-motion stimuli whereby

[2] According to this account, the onsets of both dot fields act as exogenous cues but the attentional benefit granted by the onset of the first stimulus has mostly waned by the time of the onset of the delayed dot field. The delayed dot field thus enjoys an attentional benefit that extends through the period of translation.

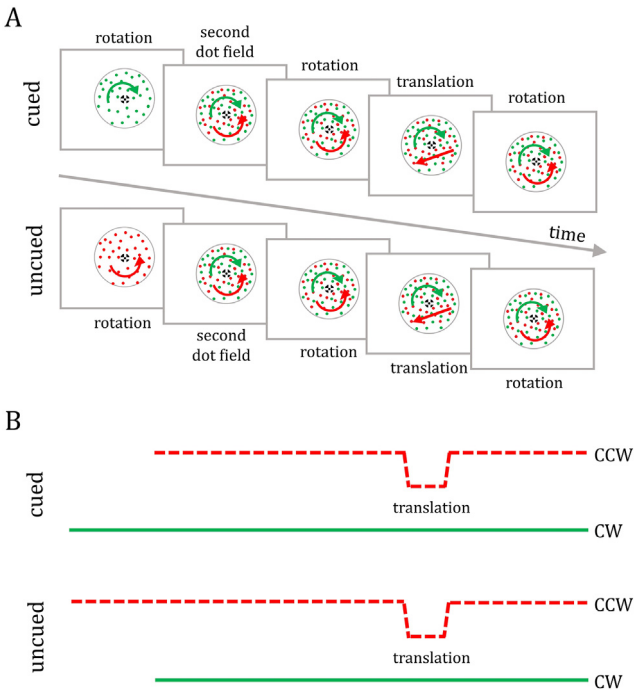

Fig. 1 – Delayed-onset design. (A) One rotating dot field appears followed by the second "delayed" dot field. Two superimposed dot fields rotate in opposite directions around the fixation point, allowing the perception of two transparent surfaces. Following the rotation, either the delayed (cued) or non-delayed (uncued) dot field translates briefly. After the translation, both dot fields continue to rotate. (B) Feature-based illustration of timeline. The two dot fields are differentiated by line style (dashed or solid) and with dot field colors indicated by the line colors. The vertical line placement indicates the different motion directions: clockwise rotation (CW), counter-clockwise rotation (CCW), and translation. The onset differences in this design result in "cued" translations occurring in the presence of the older rotation direction and "uncued" translations occurring in the presence of the newer rotation direction.

the successive motions (i.e., the rotation followed by the translation) of a cued object are preferentially processed relative to motions of an uncued object.

1.3. *Neuronal correlates of object-based attention*

In addition to providing behavioral evidence of object-based attention, variants of the original transparent-motion design have been adapted to study single-unit correlates in the non-human primate (Fallah, Stoner, & Reynolds, 2007) and neural mechanisms using functional magnetic resonance imaging (fMRI) (Ciaramitaro, Mitchell, Stoner, Reynolds, & Boynton, 2011; Ernst, Boynton, & Jazayeri, 2013). Of particular interest here is the use of these designs to study the neural substrates of object-based attention via Electroencephalography (EEG) (Khoe et al., 2005; Khoe, Mitchell, Reynolds, & Hillyard, 2008; Lopez et al., 2004; Pinilla et al., 2001; Valdes-Sosa, Bobes, et al., 1998, 2004). In the first such study, Valdes-Sosa, Bobes, et al. (1998) found that when attention was endogenously cued to one dot field, the amplitude of both P1 (134–203 msec) and N1 (244–293 msec) components elicited by an uncued translation was reduced (see also Khoe et al., 2008). Both of these components have been generally associated with extrastriate areas [i.e., beyond the level of primary visual cortex (V1)]. This early experiment used a blocked design in which subjects continuously attended to either green–or red-colored dot fields within a block of stimuli. Follow-up studies have used trial-by-trial cueing and have consistently found N1 modulation but (with the exception of Khoe et al., 2005; see below) have not found significant modulation of earlier components (Lopez et al., 2004; Pinilla et al., 2001; Valdes-Sosa et al., 2004). The source of N1 modulation in these studies was found to be consistent with the involvement of area MT+ (human middle temporal complex) and area V4, and hence associated with the changes in mid-level visual processing (Valdes-Sosa et al., 2004). Using a design introduced by Reynolds et al. (2003),[3] Khoe et al. (2005) found modulation of the N1, but, surprisingly, also found modulation of the C1 component. The C1 component is believed to reflect activities in striate cortex (i.e., area V1) (Clark, Fan, & Hillyard, 1995; Di Russo, Martinez, & Hillyard, 2003). Specifically, they reported that the posterior C1 (75–110 msec) and N1 (160–210 msec) components elicited by the translations of the cued dot field were larger than those elicited by the translations of the uncued dot field. A subsequent fMRI study, using a design similar to that of Reynolds et al. (2003), found enhanced activation in areas V1, V2, V3, V3A, and MT + for translations of a cued dot field compared to translations of the uncued dot field (Ciaramitaro et al., 2011). Overall, these studies uniformly implicate mid-level cortical areas in the performance biases seen in these stimulus designs with some studies also implicating earlier cortical areas, including area V1.

1.4. *The motion-competition explanation*

Taken together these numerous studies appear to provide abundant psychophysical and neuronal evidence of object-based attention, whereby attention embraces all the features of a cued object (e.g., a specific color and rotation direction) including new features such as a brief translation. Given that Mitchell et al. (2003) found that the cueing effect survived the removal of color differences, Stoner and Blanc (2010) observed that for this object-based account to be valid, a processing advantage would seemingly have to first be granted to the rotation direction of the cued object and that this advantage would then have to extend to the translation in an object-specific manner. The object-based account thus requires a mechanism that somehow links or "binds" successive motions (i.e., rotations and translations) with each other in an object-specific manner.

Stoner and Blanc (2010) questioned the necessity of the object-based explanation and offered an alternative *motion-competition* explanation. As reviewed in section 1.3, it has been reported that ERPs associated with the middle temporal complex (MT+) are larger in response to cued versus uncued translations in the Valdes-Sosa et al. design (Khoe et al., 2005;

[3] This was not the delayed-onset design but rather a slightly more complex design with two successive translations per trial.

Rodríguez & Valdés-Sosa, 2006; Valdes-Sosa, Bobes, et al., 1998). MT+ is thought to be the human homolog of the middle temporal (MT) and medial superior temporal (MST) areas in the macaque and is known to be involved in visual motion processing (Tootell et al., 1995; Watson et al., 1993). Stoner and Blanc (2010) noted that studies of cortical areas MT and MST have demonstrated that paired stimuli presented with the receptive fields of neurons in these areas appear to compete or inhibit one another. Thus, it has been found that superimposing a dot field moving in a neuron's anti-preferred direction upon a dot pattern moving in that neuron's preferred direction (thereby creating transparent motion) suppresses neuronal responses in area MT (Qian & Andersen, 1994; Snowden, Treue, Erickson, & Andersen, 1991). Suppressive/competitive interactions have also been reported for spatially separated stimuli in areas MT and MST (Recanzone & Wurtz, 2000; Recanzone, Wurtz, & Schwarz, 1997). In addition, Krekelberg and Albright (2005) discovered that responses to multiple stimulus components are modeled well by competitive interactions and that those interactions are not restricted to opposite directions. Specifically, they found that a model advanced to account for competitive stimulus interactions in cortical areas V2 and V4 (Reynolds, Chelazzi, & Desimone, 1999; see also; Reynolds, Pasternak, & Desimone, 2000; Williford & Maunsell, 2006) also nicely accounts for stimulus interactions within area MT. These various empirical findings are also consistent with the more recent normalization models of feature-based and spatial attention (Lee & Maunsell, 2009; Reynolds & Heeger, 2009; Ruff & Cohen, 2017; Verhoef & Maunsell, 2016).

The motion-competition explanation by Stoner and Blanc (2010) begins with the same assumption as the object-based account: the delayed-onset grants the rotation direction of the delayed dot field a momentary processing advantage (yielding larger neuronal responses) relative to the rotation direction of the dot field that appears first.[4] The motion-competition explanation diverges from the object-based account, however, in not assuming that this advantage is somehow extended to the translation in an object-specific manner. Instead, the motion-competition account simply assumes that this processing advantage persists during the translation. It follows that translations of the cued (i.e., delayed) dot field would occur in the presence of rotation direction responses that are larger than those that occur in the presence of translations of the uncued dot field. This assumption leads to the motion-competition explanation of how cueing impacts neuronal and behavioral responses in these experiments. To appreciate this explanation, first note that a key feature of the competitive stimulus interactions observed in area MT and elsewhere (and in the models that capture those interactions) is that *competing stimuli that elicit larger responses also elicit greater inhibition*. This fact coupled with the assumption of a response asymmetry to the rotations of the cued versus uncued dot fields thus leads to the prediction that responses to translations of the cued (delayed) dot field would be less inhibited (and hence larger) than responses to translations of the uncued dot field. Indeed, using simulations, Stoner and Blanc (2010) showed that a model based on competitive stimulus interactions could account for the neuronal (and, by extension, the behavioral) effects observed with the various transparent-motion paradigms reviewed above. Their motion-competition model parsimoniously accounts for these findings without the need to invoke object-specific attentional enhancement. This alternative explanation casts doubt on the object-based interpretation of the numerous studies that had used variants of the Valdes-Sosa, Cobo, and Pinilla (1998) design. As these studies seemingly constituted some of the best evidence of object-based attention (Reynolds et al., 2003), the analyses of Stoner and Blanc (2010) suggested that the case for object-based attention was less overwhelming than had been thought.

Unlike the object-based account, the motion-competition model posits that the key determinant of behavioral performance is not whether the cued or uncued object (i.e., dot field) translates but whether the translation "competes" with the rotation direction that was presented first or second regardless of which set of dots undergo those motions. To distinguish between the motion-competition and object-based models, Stoner and Blanc (2010) introduced motion (rotation direction) swaps between the two dot fields at the onset of the brief translation (see Fig. 2B). These motion swaps reverse the relationship between cueing and which rotation direction (delayed or undelayed) competes with the translation. As a consequence, the motion-competition model, contrary to the object-based account, predicts that this manipulation should reverse the performance asymmetry seen in the standard paradigm without motion swaps. Specifically, if the rotation direction is swapped at the onset of the translations, then translations of the cued (delayed) dot field should be harder to judge than translations of the uncued dot field. Stoner and Blanc (2010) found, however, that the performance advantage did not reverse with these motion direction swaps. They also found that color-swaps similarly did not reverse the performance advantage. Stoner and Blanc's (2010) findings thus supported the object-based account and ruled out the established competition and normalization models (e.g., Lee & Maunsell, 2009; Reynolds & Heeger, 2009) as an explanation for their results. Moreover, their finding that the cueing effect was dot-field specific implicated mechanisms that could distinguish between the spatially intermixed and moving dots of the two dot fields.

1.5. *Goals of the current study*

We had several goals in the current study. First, as the Stoner and Blanc (2010) study is the only study of the numerous studies cited above that provided evidence of object-based selection that cannot potentially be explained by previously identified competitive/normalization mechanisms, we wished to confirm their finding that the behavioral effect of cueing survives feature swaps. Second, since ERPs have not been previously collected with the delayed-onset design, it is

[4] The processing advantage of the delayed (newer) rotation over the older rotation could result from exogenous cueing based on the abrupt onset of the delayed dot field (Reynolds et al., 2003) and/or neuronal adaptation as suggested by Stoner and Blanc (2010). Explanations based on exogenous cueing and neuronal adaptation are not necessarily at odds as neuronal adaptation could (at least partially) underlie the ability of abrupt onsets to capture attention.

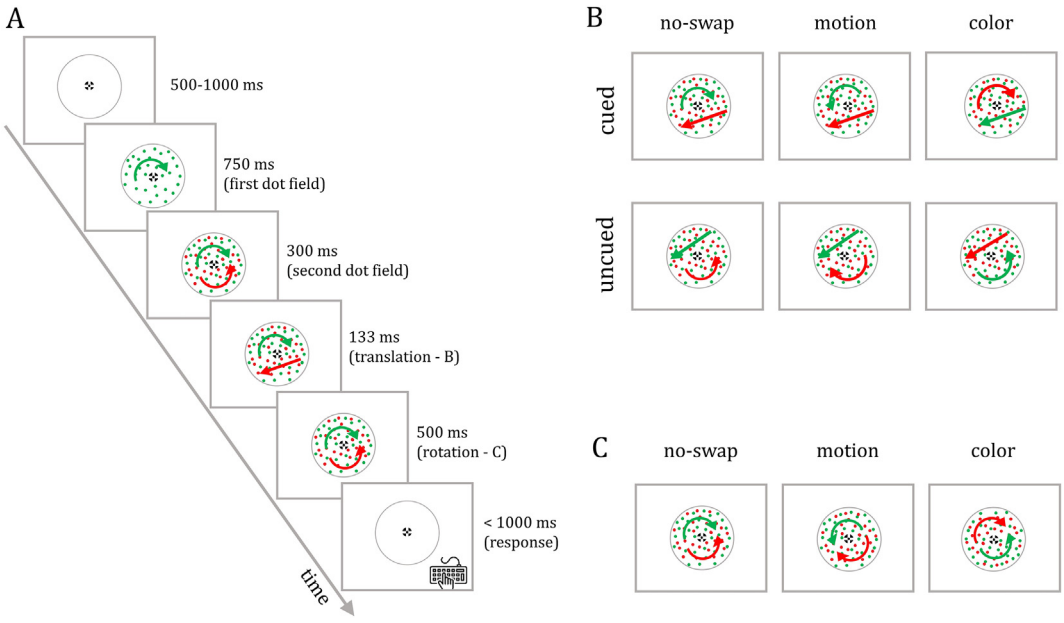

Fig. 2 – Schematic representation of stimuli. (A) Each trial started with a variable fixation period (500–1000 msec), followed by the appearance of the first dot field (i.e., surface) rotating around the fixation point for 750 msec. A second (delayed) dot field appeared and both fields rotated in opposite directions around the fixation point for 300 msec. (B) Following this period of dual rotation, one dot field translated in one of the eight directions for 133 msec. At the translation onset, the non-translating dot field either continued to rotate in its original direction ("no-swap" trials) or reversed rotation direction ("motion-swap" trials). The colors of the two dot fields were similarly swapped at the onset of translation on half of the trials ("color-swap" trials). Cued trials are defined as translations of the delayed dot field. The resulting six different conditions (2 cued/uncued x 3 feature swaps) are illustrated separately. (C) After the translation, the translating dot field either resumed its original rotation direction (e.g., "no-swap" trials) or assumed the other dot field's previous rotation direction (e.g., "motion-swap" trials), and both surfaces kept rotating for 500 msec in their newly assigned directions. The subject's response window started 100 msec after the translation onset and ended 1 sec after stimulus offset. Note that the stimulus shown in (A) is an example of the cued no-swap condition. It is a cued condition because the delayed (in this example, red) dot field translates. It is a no-swap condition as neither the colors nor rotation directions are swapped. The gray/surrounding circles highlight the circular aperture, and they were not present on the actual display.

conceivable that the ERP correlates observed using other variants of that paradigm are specific to the details of those designs (e.g., the presence of two translations per trial rather than one translation per trial as in the delayed-onset design). Indeed, as reviewed in section 1.3, different studies using slightly different designs have found somewhat different ERP results. We hypothesized, however, that the N1 modulation found in previous studies was not specific to those designs and would also be found with the delayed-onset design. This would support the conclusion by Reynolds et al. (2003) that the delayed-onset design captured the key features of the more complicated designs. Third, and most important, the ERP correlates observed in previous experiments are all subject to a motion-competition interpretation. Hence, it is unclear whether the previously identified ERPs associated with cueing are truly related to object-based attention. Thus, our goal was to identify ERPs associated with cueing that survived the feature-swaps and hence could be identified as supporting object-based attention rather than reflecting competitive interactions between direction-selective neurons. Lastly, based on their findings, Stoner and Blanc (2010) hypothesized that area V1 is involved in the object-based effects they identified. We were, therefore, interested in determining whether we might find further evidence of V1's involvement, such as seen by Khoe et al. (2005) and Ciaramitaro et al. (2011). Any such ERP modulation tentatively associated with V1 would need to survive feature swaps to be identified as supporting object-based cueing.

2. Methods

In this section, we report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study. No part of the study procedures or analyses was pre-registered in a time-stamped, institutional registry prior to the research being conducted.

2.1. *Participants*

Fifteen adult human volunteers (8 females and 7 males, age range 18–27 years) out of 20 recruited subjects completed all the experimental procedures and EEG sessions. All participants had normal or corrected-to-normal visual acuity and no

history of neurological disorders. Before their participation, they were informed about experimental procedures and signed a consent form. The data from 5 participants that failed to reach criterion performance (see section 2.3) or could not complete a full number of experimental sessions were not included. The final sample size was commensurate with previous reports (Khoe et al., 2008). The inclusion/exclusion criteria were established prior to data analyses. All procedures were carried out under the Declaration of Helsinki (World Medical Association, 2013) and approved by the local Ethics Committee of Bilkent University.

2.2. Apparatus

Visual stimuli were generated with Matlab 7.12 (The MathWorks, Natick, MA) and presented by using the PsychToolbox 3.0 (Brainard, 1997; Pelli, 1997). A 20-inch CRT monitor (Mitsubishi Diamond Pro 2070sb, 1600 × 1200 pixel resolution and 60 Hz refresh rate) was used to display the stimulus from a viewing distance of 57 cm with a chinrest to stabilize the head. All procedures were carried out in a dark room. A photometer (SpectroCAL, Cambridge Research Systems, Rochester, Kent, UK) was used for the calibration of the display. Using a digital oscilloscope (Rigol DS 10204B, GmbH, Puchheim, Germany) connected to a photodiode centered over the position of our stimuli, we continuously synchronized our EEG recordings with the stimulus/event onset times.

2.3. Training and performance testing

Before engaging in the main EEG experiment, potential participants first engaged in fixation training, the design of which also allowed identification of subjects that could not reliably fixate. We did this for several reasons. First, we wanted to ensure that subjects were not visually tracking a subset of dots (despite being instructed to neither track dots nor attend either dot field). Second, in preliminary experiments, we observed that the cueing effect was weak or absent in some subjects. We subsequently discovered that training subjects to fixate accurately, and only including subjects that could do so reliably, substantially increased the average cueing effect. The training session included central and peripheral conditions. The combination of a bull's eye and crosshair was used as a central fixation for both conditions as this fixation target has been shown to elicit reliable and stable fixation (Thaler, Schütz, Goodale, & Gegenfurtner, 2013). The bull's eye was constructed of outer and inner circles with diameters of .6° and .24°, respectively. In the central condition, across the diameter of the inner circle, a small .06° wide bar changed orientation from horizontal to vertical for 133 msec at the center of the screen. For the peripheral condition, there was an additional target (i.e., bull's eye with crosshair) located .6° away from the central target. In this condition, the bar was located inside the peripheral target rather than the central one. The fixation training started with each participant determining their threshold luminance/contrast level by adjusting the brightness of the bar till the repetitive transition was barely detectable. This was done separately for central and peripheral target conditions. After 10 repeats of this adjustment procedure for each condition (central and peripheral) separately, the estimated mean gun values at threshold detection for the two conditions were then used in the main fixation task. Since the task was personalized for each participant, the effects of differential visual acuity were minimized. The main fixation task was to detect an orientation transition (from horizontal to vertical to horizontal) in the central or peripheral target conditions. While the participants were maintaining fixation on the central fixation target, a transition in the middle of a target (either central or peripheral) occurred for 133 msec with a random onset timing during half of the trials (i.e., 108 trials of a session). Participants reported, with a key-press, whether they saw a transition or not. Under these conditions, they would not be able to notice the transition in the central condition when there was an eye movement of more than .6°. Also, they would only notice the transitions on peripheral conditions if there were a fixation break. Therefore, a high difference in detecting the transition between central and peripheral conditions suggests that the fixation was sustained. To be eligible to continue with the rest of the experiment, participants were required to have a performance value at least 25% higher than the chance level for the central condition and to have the same difference between the central and peripheral conditions. This session (2 conditions x 108 trials per condition) allowed us to train participants to sustain fixation throughout the main experiment as well as to screen their fixation performance.

Each participant that was found eligible based on fixation performance engaged in a heterochromatic flicker fusion task (Ives, 1912) to establish equiluminance between the red and green guns of the monitor with a flicker rate of 60 Hz with the goal being to have red and green dot fields of approximately equal salience. Using a 2 × 2° square stimulus, the red gun was held constant at maximum intensity (19.7 cd/m^2) and the green gun was adjusted until a minimal flicker was reported. This procedure was repeated 10 times. The averaged green value was then used for each participant in preliminary practice sessions as well as in the main EEG experiment. To ensure that participants understood the task completely and were able to achieve above-chance level performance, each participant first engaged in a practice session. The practice session consisted of the basic cued and uncued conditions without feature-swaps (no-swap conditions, see section 2.4). There were 240 trials per condition, leading to a total of 480 trials (2 cued/uncued conditions x 240 trials). The observers were considered to have achieved criterion performance during the practice session if they correctly indicated the translation direction for more than 25% of the total trials (i.e., more than 120 trials). This criterion performance level corresponds to twice the chance level (12.5% chance level based on 8 different translation directions, see section 2.4).

2.4. Stimuli and procedure

Visual stimuli consisted of two superimposed circular fields (3.3° diameter) of randomly distributed dots rotating in opposite directions around a central fixation target on a black background (.16 cd/m^2). As in fixation training, we used a fixation target that was a combination of a bull's eye and crosshair, which has been shown to elicit reliable and stable fixation (Thaler et al., 2013). The diameters of the inner and

outer circles of the bull's eye were .24° and .6°, respectively. The average density of each dot field was 5 dots per square degree of visual angle. Each dot had a diameter of .05°. The two dot fields rotated (in opposite directions) with a speed of 81°/sec around the fixation target. One dot field was red and the other was green, the luminance values were equiluminant based on the heterochromatic flicker fusion task (see above). Whether the red or green dot field appeared first was randomized across trials.

Subjects initiated a trial with a key-press. To ensure that subjects had time to stabilize fixation, trials started with the fixation target (Fig. 2A). After a variable duration of between 500 msec and 1 sec, one dot field appeared, which rotated (clockwise or counter-clockwise) continuously around the fixation target for 750 msec. Next, the second ("delayed-onset") dot field appeared and rotated in the opposite direction (clockwise or counter-clockwise). After the onset of the second dot field, both dot fields continued to rotate for 300 msec. Following this 300 msec period, either the delayed (a "cued translation") or the non-delayed (an "uncued translation") dot field translated briefly (133 msec) in one of eight (cardinal and diagonal) directions. Participants reported the translation direction via the numeric keypad of keyboard. They were instructed to think of the '5' key as the center and map the eight directions to the remaining keys, similar to a compass. Our design also included motion or color swaps on some trials yielding six different conditions (2 cued/uncued x 3 feature swaps; Fig. 2B). At the onset of the translation, the non-translating dot field either continued to rotate in its original direction ("no-swap" trials) or reversed rotation direction ("motion-swap" trials). After the translation, the translating dot field either resumed its original rotation direction ("no-swap" trials) or assumed the other dot field's previous rotation direction ("motion-swap" trials). Similarly, the colors of the two dot fields were swapped at the onset of translation ("color-swap" trials). Motion and color swaps (when introduced) persisted until the end of the trial. The rotation duration after the translation offset was 500 msec (Fig. 2C). The percentage of coherently moving dots was 60% and all dots translated at a speed of 2.26°/sec. The remaining dots were distributed equally to move in the other seven directions. The translation duration was kept constant at 133 msec. Different random-dot fields were used for each translation direction, color-order (i.e., red or green dot field was presented first in a trial), and experimental conditions. The demonstrations for the main experimental conditions (Fig. 2B) are available at Open Science Framework (https://osf.io/kpjgv/).

Response registration and timing were based on the translation onset. If the participant responded earlier than 100 msec after the translation onset or failed to respond until 1 sec after the stimulus left the screen, the trial was repeated. The participants either reported the direction or skipped the trial in case of an interruption (e.g., fixation break). Skipped trials were also repeated. The response registration continued until the end of a trial or the participant passed forward to the next trial. Prior to each session, all participants were informed about the limited response time and told that the trials would be repeated if they failed to respond within this time range. The participants were informed that the translating dots could be of either field and that only a subset of one of the dot fields translated coherently. Accordingly, they were told to attend to the entirety of both dot fields to maximize their ability to discriminate the global direction of those translations. Hence, there was no incentive to selectively attend to one of the two dot fields or any subset of dots. As described above, each participant had been trained to fixate and screened based on their ability to do so. Subjects were instructed to fixate throughout each trial. Each participant completed a main experimental session including 96 trials per condition.

2.5. EEG data acquisition and preprocessing

EEG recording and preprocessing steps were similar to those described previously (Akyuz, Pavan, Kaya, & Kafaligonul, 2020; Kaya & Kafaligonul, 2019). A 64-channel MR-compatible system (Brain Products GmbH, Gilching, Germany) was used to record high-density EEG activities. Prior to each experimental session, we carefully placed the EEG cap on a participant's head. The placement of scalp electrodes was based on the extended 10–20 system. Two electrodes were used as reference (FCz) and ground (AFz). We used q-tips and a syringe with a blunt tip to apply conductive paste (ABRALYT 2000 FMS, Herrsching–Breitbrunn, Germany) in order to reduce the impedance of each electrode below 10 kΩ. The impedance levels were monitored during the sessions for reliable recording. EEG signals were sampled at 5 kHz and band-pass filtered between .016 and 250 Hz. We stored the EEG data, event markers, and behavioral responses using the Vision Recorder Software (Brain Products GmbH, Gilching, Germany) for offline analyses.

EEG preprocessing steps were carried out offline with Brain Vision Analyzer 2.0 software (Brain Products GmbH, Germany). First, EEG signals were down-sampled to 500 Hz and filtered using a zero-phase Butterworth band-pass filter (.5–100 Hz, 24 dB/octave) and a 50 Hz notch filter (50 Hz ± 2.5 Hz, 16th order). Using the recorded signal from the electrocardiogram electrode, the cardioballistic artifacts were removed (Allen, Polizzi, Krakow, Fish, & Lemieux, 1998). After these preprocessing steps, the data were next divided into epochs starting from 500 msec before the onset of the first dot field to 1 sec after the translation offset. Independent component analysis with the Infomax algorithm was used on the data to remove common EEG artifacts such as eye blinks. Lastly, we used a combination of manual and automated selection to find and remove trials contaminated with oscillations over 50 μV/msec, voltage changes more than 200 μV in 200 msec, or changes of less than .5 μV in 100 msec window. Bad channels were restored using spherical spline interpolation (Perrin, Pernier, Bertrand, & Echallier, 1989). After applying these preprocessing steps, on average 92% of trials (SEM = 1.11%) were preserved for further ERP analyses. The excluded trials during the EEG preprocessing stage were also not used in the analysis of the behavioral data.

2.6. ERP analyses

After preprocessing, we averaged the EEG signals from each electrode across all valid trials to compute ERPs time-locked to the onset of the first rotating dot field. For further smoothing,

these averaged ERPs were filtered with a low-pass filter (6th order zero-phase Butterworth IIR filter with 40 Hz cut-off frequency). A baseline correction was also applied using the 100 msec time window before the onset of the first dot field. We adopted a two-step approach in our ERP analyses. The current study is the first EEG investigation using the delayed-onset design, and the first to incorporate the feature-swaps that allow identification of object-based mechanisms. Given these key design differences between the current and previous studies, we could not assume that we would see the ERPs previously identified in those studies. For this reason, we used a data-driven approach at the first stage of the analyses to comprehensively evaluate modulations in the spatiotemporal profile of neural activity. We performed the cluster-based permutation test integrated into the Fieldtrip toolbox to identify spatiotemporal clusters associated with the significant modulations of ERPs (Oostenveld, Fries, Maris, & Schoffelen, 2011). This is a data-driven non-parametric framework to overcome multiple statistical comparisons (Type I error) and to cluster selected samples objectively (Groppe, Urbach, & Kutas, 2011; Maris & Oostenveld, 2007). Our experimental design included conditions with a sequence of visual events during each trial. It is possible to have overlapping components and leakage from the pre-translation period to the evoked activities after translation onset. On the other hand, the types of visual events were the same up to translation onset for the cued and uncued conditions. Therefore, we compared the cued trials with the uncued trials (i.e., baseline condition) to avoid any potential confounding factor. Our first goal was to identify spatiotemporal ERP clusters that reflected the effects of cueing across all of the swap conditions. Accordingly, we combined all the waveforms (i.e., averaged across swap conditions) for the cued and uncued conditions, separately. These two combined waveforms (cued *vs* uncued) were compared at each electrode location and time point (2-msec bin) using a paired samples *t*-test. The samples with *t-values* exceeding an uncorrected alpha level of .05 were clustered together based on spatial (i.e., electrode location) and temporal (i.e., time point) proximity. At least three neighboring electrodes were required to form a cluster. The *t-values* within a cluster were summed up to have cluster-level statistics. To obtain a null hypothesis distribution of the cluster-level statistics, 10,000 random permutations of the original data were generated using the Monte Carlo method. A cluster in the experimental data was considered to be significant when it fell in the highest or the lowest 2.5th percentile of the generated distribution (corresponding to the significance level of a two-tailed test). That is to say, since the tests were two-sided, the significance threshold for testing the null hypothesis (alpha level) was .025 (Maris & Oostenveld, 2007). Previous EEG studies identified the evoked activities and components within the first 400–500 msec time window after the translation onset (e.g., Khoe et al., 2005, 2008; Pinilla et al., 2001). By following the same approach and conventions, we restricted our analyses (i.e., cluster-based permutation test) to the first 400 msec time range and hence identified the modulations.

Following the cluster-based permutation test, we identified the time-range and electrode locations (i.e., exemplar sites) of spatiotemporal clusters associated with the significant effect of cueing. Using the exemplar sites, we displayed the time-courses of evoked brain activities to all six conditions (2 cued/uncued x 3 swap conditions) for illustrative purposes. At the second stage of ERP analyses, we averaged activities within the identified time-range of a cluster and carried out a regression analysis between these mean potentials and behavioral performance values across all six conditions to determine how these potentials related to performance. The correlation and association between the mean potentials and performance measures across the six experimental conditions were evaluated through linear regression fits having intercept and slope as coefficients. Next, to elucidate the source of ERP modulations, we computed the peak amplitudes and latencies of ERP components over the exemplar sites. We performed a two-way repeated-measures ANOVA (with cueing and swapping as factors) on these ERP metrics. When Mauchly's test indicated that the assumption of sphericity had been violated, the Greenhouse–Geisser correction was applied. The corresponding epsilon (ε) values (i.e., sphericity estimates) are supplied when the ANOVA results are presented. It is important to note that the cluster-level statistics were built upon *t-values* from selected samples and the permutation test outcome does not directly provide any additional metric for the effect sizes. These additional ANOVAs thus allow estimation of effect sizes for all experimental factors as well as any two-way interaction.

3. Results

3.1. *Behavioral results*

The average performance values of each participant and group-averaged data are shown in Fig. 3. All participants performed above the chance level (12.5% based on 8 translation directions). A two-way repeated-measures ANOVA with

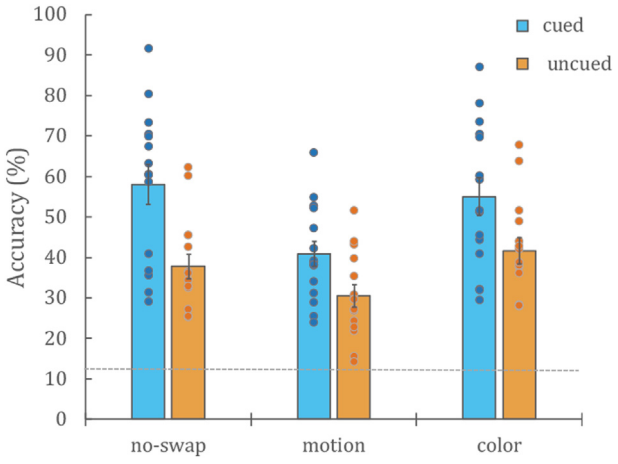

Fig. 3 – Behavioral results (n = 15). The percentage of correct responses for each condition with blue and orange bars corresponding to cued and uncued conditions, respectively. Dots represent the data from individual participants. The dashed line indicates chance performance level (i.e., 12.5%). Error bars ± SEM.

cueing (cued *vs* uncued) and swapping (no-swap, motion, color) as factors indicated significant main effects for both cueing ($F_{1,14} = 49.545$, $p < .001$, $\eta_p^2 = .780$) and swapping ($F_{1.18,16.52} = 8.676$, $p = .007$,$\eta_p^2 = .383$, $\varepsilon = .59$).[5] In agreement with Stoner and Blanc (2010), we found that observers were significantly better at judging translations of the cued (delayed) dot field (M = 51.34%, SEM = 3.46%) than of the uncued (non-delayed) dot field (M = 36.66%, SEM = 2.39%) in both the presence and absence of swaps. However, the overall accuracies of observers were reduced for the trials that included motion swaps (see also Supplementary Table S1). Moreover, a significant two-way interaction between cueing and swapping was revealed ($F_{2,\ 28} = 4.971$, $p = .014$, $\eta_p^2 = .262$). In summary, while the cueing advantage survived motion swaps, the inclusion of motion swaps not only decreased the overall accuracy but also decreased the accuracy difference between cued and uncued conditions. To further understand the differential effects of cueing, we calculated the difference performance values between cued and uncued conditions for all swap conditions (Table 1, two right-most columns for descriptive statistics) and performed paired *t*-tests on these difference values. Bonferroni-corrected pairwise comparisons indicated that the difference performance values of conditions with motion swaps were significantly different from that of the no-swap condition ($t_{14} = 3.225$, $p = .018$, Cohen's $d = .833$). Additionally, there was also a significant difference between the cueing effects of no-swap and color swap conditions ($t_{14} = 2.727$, $p = .049$, Cohen's $d = .704$). Despite the impact of feature swaps, the accuracy for the cued trials was significantly greater than for uncued trials for all conditions (Table 1).

3.2. EEG results

In line with behavioral results, we found a robust difference in the evoked activities elicited by the cued conditions compared to the uncued. A cluster-based permutation test on the averaged ERPs in the 400 msec time-range after translation revealed a spatiotemporal cluster associated with the significant effect of cueing (cued *vs* uncued, cluster-level $t_{sum} = -7437$, $p = .002$). This cluster was mainly within the 238–326 msec time window. While the cluster spread over some centro-parietal electrodes, it was mainly over occipital and parieto-occipital scalp sites (Fig. 4). Over these electrodes, the average activity for cued trials was larger in amplitude in a negative direction (i.e., more negativity) compared to the uncued trials.

The cluster-based permutation test also indicated an earlier (134–224 msec) and nonsignificant cluster (cued *vs* uncued, cluster-level $t_{sum} = -2035$, $p = .042$). The average magnitude of cueing effect was smaller for this early time-range. In other words, the analyses reported some electrodes which were differentially activated for at least 20 msec such as medial parieto-occipital and parietal scalp sites (Supplementary Fig. S1). However, these observed cueing effects did not survive at the cluster level statistics which was based on the spatiotemporal domain. To elucidate the differential effects of swapping on the spatiotemporal profile of the neural activity, we subtracted the averaged ERPs of the cued trials from those of uncued trials for each swap condition. Based on the observed impact of motion and color swaps on behavioral performance, we were particularly interested in comparing the difference waveforms of these conditions with those of no-swap condition. These and other similar comparisons did not reveal any additional and meaningful spatiotemporal cluster.

Electrodes that were part of the identified significant cluster for more than 70 msec (i.e., more than 75% of the 238–326 msec time-range) were selected as exemplar sites (Fig. 4). The averaged potentials over these electrodes are shown in Fig. 5A. In these averages, there were robust evoked activities peaking around 300 msec (N1 component) post translation. The identified time window of the cluster corresponded to the range of this negative N1 component. In this component range, the cued conditions elicited larger amplitudes compared to the uncued trials (Fig. 5B). This was also reflected in the mean values within the identified time window (238–326 msec) for each swap condition (Fig. 5C). We performed a linear regression fit using both of these mean potentials and performance values. The analyses revealed a significant association between these measures across different conditions ($R^2_{adj} = .577$, $p < .05$; Fig. 6).

To further understand the nature of these ERP modulations, we performed repeated-measures ANOVAs with cueing and swapping as factors on the peak amplitudes and latencies of the N1 component (Fig. 5A and B). The peak amplitude was significantly dependent on cueing ($F_{1,14} = 20.549$, $p < .001$, $\eta_p^2 = .595$). Neither the main effect of swapping ($F_{2,\ 28} = .179$, $p = .837$, $\eta_p^2 = .013$) nor the two-way interaction between cueing and swapping ($F_{2,\ 28} = .176$, $p = .840$, $\eta_p^2 = .012$) were significant. Similarly, a secondary ANOVA on the mean potentials (Fig. 5C) only revealed a significant main effect of cueing and did not indicate a significant main effect of swapping or a two-way interaction. ANOVA applied to the peak latencies did not reveal significant main effects (cueing: $F_{1,\ 14} = .084$, $p = .777$, $\eta_p^2 = .006$; swapping: $F_{1.33,\ 18.65} = 2.438$, $p = .129$, $\eta_p^2 = .148$, $\varepsilon = .67$) or two-way interaction ($F_{2,\ 28} = 1.862$, $p = .174$, $\eta_p^2 = .117$). Although the above analysis revealed no significant impact of cueing on peak latencies, the time

Table 1 – The results of post-hoc paired samples *t*-tests for the cueing effect (cued *vs* uncued) for each swap condition. The descriptive statistics (mean, SEM) are based on the difference between cued and uncued conditions. Significant *p* values (Bonferroni-corrected $p < .05$) are highlighted in bold.

	t_{14}	*p*	Cohen's d	Mean (%)	SEM (%)
no-swap	6.626	**< .001**	1.711	20.217	3.051
motion	5.096	**< .001**	1.316	10.426	2.046
color	4.308	**< .001**	1.112	13.415	3.114

[5] Mauchly's test indicated that the assumption of sphericity had been violated for main effect of swapping, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity.

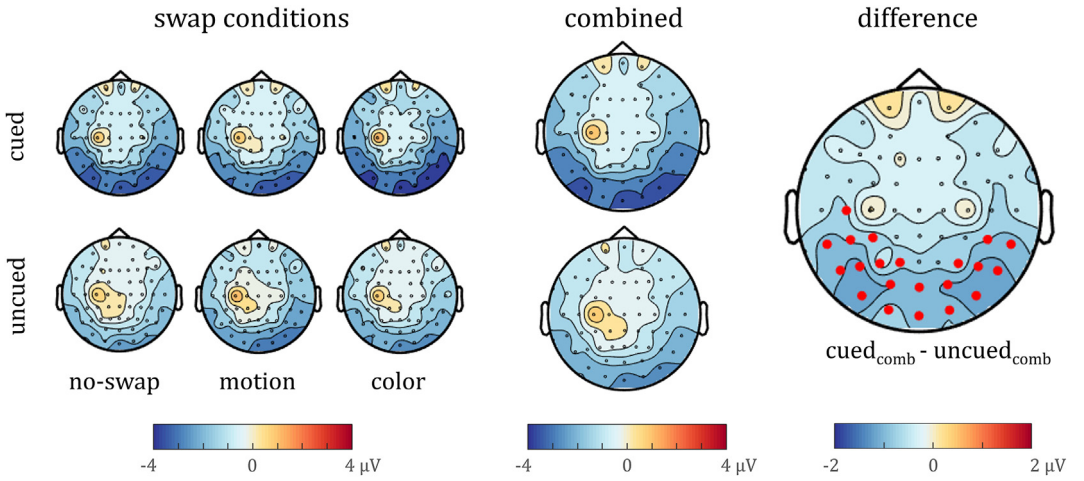

Fig. 4 – Voltage topographical maps of the averaged waveforms within the identified time window (238–326 msec). The voltage topographical maps of cued and uncued conditions are shown in separate rows. The averaged activities of each swap condition, combined waveforms across swap conditions, and the difference between them (cued$_{comb}$-uncued$_{comb}$) are displayed on the maps in separate columns. The result of the cluster-based base permutation test comparing the combined waveforms (cued$_{comb}$ *vs* uncued$_{comb}$) is indicated in the last column. The electrodes that were part of the significant spatiotemporal cluster for at least 70 msec (i.e., more than 75% of the time-range) were chosen as exemplar electrodes and are marked by red-filled circles on the right-most topographical map (i.e., Oz, O1, O2, POz, PO3, PO4, PO7, PO8, P1, P3, P4, P5, P6, P7, P8, CP3, CP5, CP6, C5, TP8, TP7).

courses of an ERP component elicited by two different conditions can differ without having different peak latencies (Luck, 2005). Indeed, an inspection of Fig. 5B suggests that the evoked activities of the cued conditions might have had an earlier onset than those of the uncued conditions. To determine whether these onset latencies differed significantly, we computed the *latency to the half-peak amplitude* applying methods used in previous neurophysiological studies (Kafaligonul, Albright, & Stoner, 2018; Sundberg, Mitchell, Gawne, & Reynolds, 2012). This latency measure was defined as the time point when the response reached half the difference between the peak amplitude and baseline amplitude. We established this latency by first identifying the first 5 successive 2-msec bins within the N1 time-range that exceeded this midpoint estimate. The latency was taken to be the first of these bins. We computed this metric for all conditions of each participant. A two-way repeated-measures ANOVA did not reveal any main effect of cueing ($F_{1,\ 14} = .011$, $p = .919$, $\eta_p^2 = .001$), swapping ($F_{2,\ 28} = 2.809$, $p = .077$, $\eta_p^2 = .167$) or a significant two-way interaction ($F_{2,\ 28} = .162$, $p = .851$, $\eta_p^2 = .011$). Thus, neither analysis of the peak latency nor the latency to the half-peak amplitude revealed a significant effect of cueing or swapping on the time course of the N1.

These analyses thus demonstrated that the N1 amplitude was impacted by cueing but revealed no significant effect of swapping thereby supporting an object-based account of our results and providing no support for feature-based effects (such as those proposed by the motion-competition explanations). To further investigate the possibility that swapping (and hence feature-based effects) might nevertheless have impacted the N1, we applied Bayesian statistics to all the N1 component metrics (peak amplitude, latency, and half-peak latency). Consistent with our ANOVA results, a Bayesian repeated-measures ANOVA suggested that the changes in peak amplitudes were best represented by the model including only cueing as the main factor, compared to the null model ($BF_{10} = 132.3$). With regards to the peak and half-peak latency values, the outcome of Bayesian tests mainly supported the null hypotheses similar to the previously reported ANOVA results (see Supplementary Table S2). In conclusion, the analyses on the N1 component revealed a robust cueing effect on amplitude, with N1 responses to translations being larger (i.e., more negative) for cued conditions than for uncued conditions. This trend held for all swap conditions. Since motion-competition and other feature-based mechanisms predict that feature swaps should reverse the cueing effect (Stoner & Blanc, 2010), our results rule out those explanations for the N1 effect observed in the current study. Conversely, we observed no significant impact on the time course of the N1 for either our cueing or feature-swapping manipulations. These follow-up tests further emphasize that the strength modulations (rather than pure latency shifts) mainly contributed to the observed cueing effect on the N1 component (Fig. 5B) over the identified occipital and parieto-occipital scalp sites.

4. Discussion

4.1. *Summary*

The present study confirmed earlier findings (Reynolds et al., 2003; Stoner & Blanc, 2010) that the delayed onset of one of two superimposed counter-rotating dot fields yields an advantage in judging the direction of subsequent brief

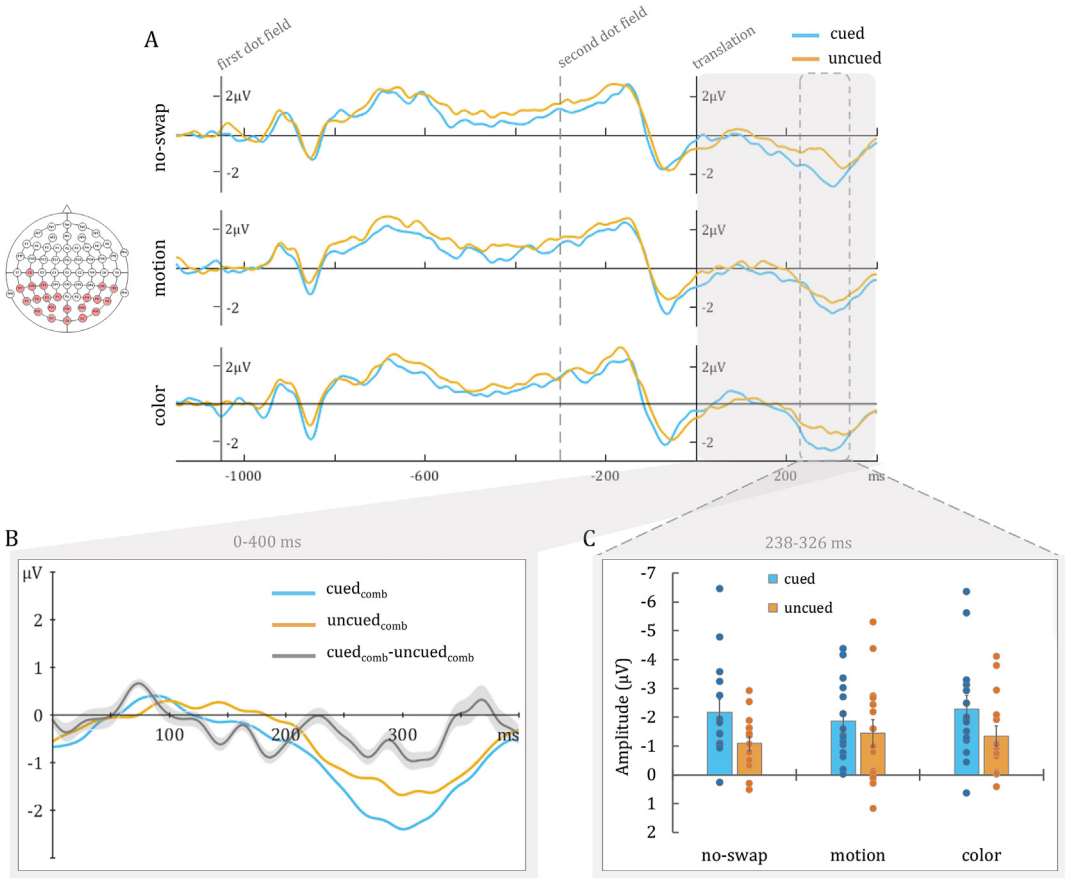

Fig. 5 – Averaged activities and derived waveforms from the exemplar scalp sites (n = 15). The locations of exemplar electrodes that were used in analyses are shown on a head model (i.e., Oz, O1, O2, POz, PO3, PO4, PO7, PO8, P1, P3, P4, P5, P6, P7, P8, CP3, CP5, CP6, C5, TP8, TP7). (A) The averaged ERPs for each swap condition are displayed in separate plots. In each plot, the blue and orange curves correspond to evoked activities for cued and uncued conditions, respectively. The ERPs were time-locked to the onset of the translation and displayed in the range from the start of each trial to 400 msec after translation. The dashed lines indicate the appearance of the first (−1050 msec) and second dot fields (300 msec). The 100 msec time window before the onset of the first dot field (−1150 msec, −1050 msec) was used as the baseline period. The identified time window based on the cluster-based permutation test is highlighted by a dashed rectangle. (B) The averaged combined waveform for the cued (blue) and uncued (orange) conditions after translation. The difference waveform ($cued_{comb}$-$uncued_{comb}$) is indicated by the thick gray curve, and the shaded area corresponds to the standard error (±SEM) across participants. (C) Bar plots displaying the averaged amplitudes within the identified time window (238–326 msec) for each condition (2 cued/uncued x 3 swap conditions). The blue and orange bars correspond to cued and uncued conditions, respectively. The dots represent the data from individual participants. Error bars ± SEM.

translations of the delayed dot field relative to translations of the non-delayed dot field. By introducing motion-direction and color "feature swaps" simultaneously with the translation onset, we also confirmed Stoner and Blanc's (2010) finding that this performance advantage is indeed specific to the delayed dot field rather than to the overall color and/or motion direction configuration of these stimuli. Most importantly, we found ERP correlates of this processing advantage in the N1 component range. The spatial distribution of this ERP component was consistent with the involvement of mid- and perhaps low-level visual areas. Our findings rule out explanations based on feature-specific selection as well as those based on feature competition/normalization. They support an object-based interpretation of the cueing effects observed in the delayed-onset transparent-motion design used here, as well as in the studies that have used variations of the original Valdes-Sosa, Cobo, and Pinilla (1998) two-translation design (Fallah et al., 2007; Khoe et al., 2005, 2008; Lopez et al., 2004; Mitchell et al., 2003, 2004; Pinilla et al., 2001; Reynolds et al., 2003; Rodriguez et al., 2002; Wanning et al., 2007). Our results highlight the need for a model of object-based attention that can account for the ability to discriminate between the densely intermixed elements of two spatially superimposed and rapidly moving objects.

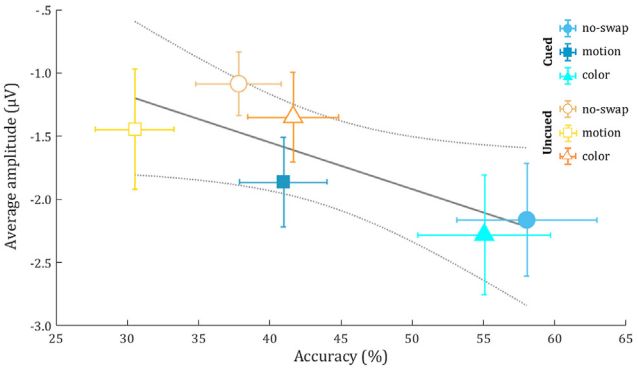

Fig. 6 – Averaged potentials in the time-range of the identified cluster (238–326 msec) with the performance values for each condition (2 cued/uncued x 3 feature swaps). The locations of exemplar electrodes used in our analyses are shown in Fig. 5. Filled and open symbols correspond to the cued and uncued conditions, respectively. Each swap condition is represented by different symbols. Vertical and horizontal error bars correspond to the variance across observers (±SEM). The black solid line indicates the best linear fit and dotted lines denote the 95% confidence intervals on the linear fit.

4.2. *Object-based versus motion-competition explanations*

As reviewed in section 1.4, Stoner and Blanc (2010) offered a motion-competition account that parsimoniously accounted for previous findings without the need to invoke object-specific attentional enhancement. Critically, the motion-competition account asserts that the key determinant of behavioral and neuronal responses is whether the translation competes with a new or old rotation irrespective of which objects (i.e., dot fields) happen to undergo those motions. To test that prediction, Stoner and Blanc (2010) introduced the motion-swap manipulations used in the current study. If the motion-competition account were valid, this manipulation should reverse the cueing effect: translations of the undelayed ("uncued") dot field should yield better accuracy than translations of the delayed ("cued") dot field. Contrary to that prediction, they found (and we confirmed in the current study) that the performance advantage conferred by delayed onset was not reversed by motion swaps. These results thus ruled out the motion-competition explanation and demonstrated that the advantage was indeed object-specific. Most importantly, these findings demonstrated that the link between rotations and translations are the spatially intermixed dots that undergo those motions. As the receptive fields of neurons in areas MT and MST (medial superior temporal area) are too large to distinguish the dots from different fields, Stoner and Blanc (2010) suggested that earlier areas such as V1 (and/or V2) were likely involved and outlined a mechanism involving top-down feature-specific interactions between areas MT (and/or MST) and area V1 coupled with spatially local cooperative interactions between neurons tuned to different features within V1. The current study was motivated, in part, to look for further evidence of that mechanism.

4.3. *Behavioral findings*

Our behavioral findings confirmed the dot specificity of the cueing effect found by Stoner and Blanc (2010): behavioral accuracy was greater for translations of the cued (delayed) dot field than for translations of the uncued dot field even in the presence of motion and color swaps. Consistent with the results of Stoner and Blanc (2010), we did, however, find a slight (but significant) decrease in the cueing effect when introducing color swaps (see section 4.4). Our results did diverge slightly from those of Stoner and Blanc (2010). Specifically, we found that motion swaps resulted in an overall decrease in performance (i.e., for both cued and uncued trials) as well as a decrease in the cueing effect. They found neither of these effects. The explanation for these discrepancies is unclear but may reflect small differences in the two designs. For example, in the current study, the translation duration (133 msec) was considerably longer than that (40 msec) used by Stoner and Blanc (2010). This longer duration was intended to yield more reliable behavioral responses but may have also resulted in the observed discrepancies: the dot-field specificity of the cueing effect may not survive the full 133 msec. In addition, the dot fields in Stoner and Blanc (2010) were significantly brighter than those used in the current study and this difference might conceivably account for the differential impact of motion swaps and/or color swaps in the two studies. How these brightness differences might be related to the different findings is not immediately clear but might be related to the differential attentional salience accompanying swaps. Lastly, it should be emphasized that the differences in the findings of the two studies are relatively small and may simply be due to chance variations in the subject pool: different subjects likely have slightly different mechanisms or strategies underlying their performance in these tasks. Further research, including the implementation of models (see section 4.7), may shed light on the small differences in the findings of the two studies.

4.4. *Role of color*

The two dot fields differed in color in the original Valdes-Sosa, Cobo, and Pinilla (1998) design as well as in most subsequent studies using variants of that design and hence color could conceivably have mediated the cueing effects found in those studies. As previously noted, however, Mitchell et al. (2003) found that the cueing effect survived the removal of color differences in the two-translation design. The results of the color swap experiments documented here, and by Stoner and Blanc (2010), demonstrate that the cueing effect is mostly not color specific: swapping the colors of the two dot fields did not reverse the cueing effect. It did, however, reduce the cueing effect very slightly. The significance of this slight reduction is not immediately clear but modeling studies like those referenced in section 4.6 may shed light on these findings. While these findings do not reveal a substantial role for color in the preferential processing of motion documented in those studies. They do not mean, however, that the color of the cued dot field is not itself granted a processing advantage when color differences are present. Indeed, Fallah et al. (2007), using the delayed-onset design, found a color processing advantage in the responses of color-selective neurons in area V4. Those

findings revealed that the color, as well as the motion direction of the delayed dot fields as revealed in other studies, enjoy a processing advantage as would be predicted from the object-based account. We predict that this processing advantage would also be dot-field specific (rather than color-specific) and hence extend to a "new" color as occurs in our color-swap conditions.

4.5. *Spatial attention*

The finding that the cueing effect is specific to the individual dots might suggest that, despite the spatial superimposition of the two dot fields, subjects in these experiments were attentively tracking a subset of these dots thus employing a type of spatial attention (i.e., with a moveable "spotlight"). We think this is unlikely for several reasons. First, there is no motivation for subjects to attend one set of dots or the other as the translations of the two dot fields occur with equal probability. Second, the subjects were firmly and explicitly instructed to diffusely attend and not attend to either dot field or a subset of the dots. Third, only a subset of the dots translated coherently so, as the subjects were informed, attending to the global motion is a better strategy than attending to a particular dot or set of dots. Finally, Intriligator and Cavanagh (2001) found that the spatial resolution of attentive tracking is much coarser than visual resolution. Thus, while the dots of the two dot fields in our study can be visually resolved, the close spatial proximity of the spatially intermixed rapidly moving dots would seemingly defeat any attempt to attentively track them.

4.6. *ERP findings*

Using a two-translation design in which the first translation serves as an exogenous cue, Khoe et al. (2005) found evidence that exogenous cueing modulated the early C1 component (75–110 msec) as well as the N1 component (160–210 msec). The C1 effect is consistent with the involvement of early areas, including area V1, whereas the N1 component is generally thought to involve mid-level cortical areas. Earlier ERP studies using the two-translation design had included an endogenous cue (fixation point color) and did not find a C1 effect (Lopez et al., 2004; Pinilla et al., 2001). Moreover, while those studies reported N1 effects, the N1 had a later time course (244–293 msec) than that of Khoe et al. (2005). In the present study, the ERP modulations were similarly most prominent in the N1 component range (238–326 msec). More importantly, these modulations indicated dot-field specific cueing that survived feature swaps (i.e., of motion and/or color). Compared to the prior EEG findings using transparent-motion (e.g., Khoe et al., 2005, 2008; Lopez et al., 2004; Pinilla et al., 2001; Valdes-Sosa, Bobes, et al., 1998; Valdes-Sosa et al., 2004), the peak of N1 component observed in this study was somewhat delayed and/or the overall component range was extended in time. One potential explanation for these differences lies in the differing rotation speeds used in the studies. It has been established that the amplitude and latency of the N1 component elicited by a motion-onset can be reduced and delayed by a previous period of motion stimulation (Müller, Göpfert, Breuer, & Greenlee, 1998; Pinilla et al., 2001). More to the point, using a transparent-motion design similar to the one here, Pinilla et al. (2001) demonstrated that N1 peak latency to brief translations is dependent on the preceding rotation speeds of the two dot fields. They found that the peak latency of the N1 was shortest when the dot fields were stationary prior to the translation and increased as the speed of the background rotation was increased. Our rotation speed of 81°/s was greater than those used in previous EEG studies and based on the findings of Pinilla et al. (2001), this is consistent with a delay of several tens of milliseconds. Pinilla et al. (2001) found, moreover, that the duration of this component increases as the rotation speed is increased. Consistent with their findings, the latency and duration of the N1 reported here was similar to that reported by Pinilla et al. (2001) for the highest rotation speed (i.e., 80°/sec). Lastly, our translation was of longer duration and this could plausibly contribute to both a delay in the peak of the N1 and longer duration relative to the findings of Pinilla et al. (2001).

4.7. *What is the neuronal basis of object-based attention?*

As discussed, our findings appear to rule out feature-based (i.e., direction-of-motion) normalization/competition models as the explanation of the cueing effect observed in this transparent-motion design. More generally, Stoner and Blanc (2010) have argued that the dot-field specificity of the cueing effect appears to rule out mechanisms that rely solely on mid-level cortical areas (e.g., area MT), as receptive fields in those areas are too large to distinguish between the intermixed dots of the superimposed dot fields. Stoner and Blanc (2010) noted that area V1 receptive fields, on the other hand, are small enough to contain mostly dots of one field or the other at any given moment in time. Based on these observations, they outlined a model involving "cooperative" (recursive excitatory) connections within direction-of-motion V1 hypercolumns[6] (Hubel & Wiesel, 1974; Ts'o, Zarella, & Burkitt, 2009). Stoner (2010) implemented a simple neuronal network version of this model and demonstrated that it could replicate the dot-field specific cueing observed by Stoner and Blanc (2010). Stoner (2010) also demonstrated that this model can be readily be extended to include hypercolumns tuned to additional feature dimensions (e.g., color) and could thus account for the spreading of attention across feature dimensions (e.g., Ernst et al., 2013; Schoenfeld et al., 2003; Schoenfeld, Hopf, Merkel, Heinze, & Hillyard, 2014; Wanning et al., 2007). Because this model achieves object-specific spreading of an attentional bias by relying on the fine-grained locations of object parts (e.g. texture elements or edges) and occurs before the feature dimensions of color and motion are partially segregated in higher-order areas, it sidesteps the need to postulate a higher-order mechanism that would somehow identify which features belong to which objects and then coordinate the activity

[6] Hypercolumns are defined here as a complete set of cortical columns (or just neurons) that encompass the whole range of a given variable (e.g., direction-of-motion) and which have approximately the same receptive field retinotopy.

in the separate cortical areas (such as MT and V4) specialized for different feature dimensions (Duncan, 1996).

While this model has the appeal of offering a concrete and simple means to account for the object-based effects described in this and related studies, neither our EEG findings nor earlier studies have demonstrated ERP modulations that definitively arise from area V1 (though see Khoe et al., 2005). To date, the best direct evidence of area V1's involvement in object-based attention appears to come from imaging studies (Ciaramitaro et al., 2011; Ernst et al., 2013). We think it is likely that EEG recordings have insufficient spatial resolution to test our proposal of area V1 and/or V2's involvement and that single-cell recording, such as used to examine area MT's involvement in surface-based attention (Wanning et al., 2007), will be needed to resolve the roles of these lower-order cortical areas.

Lastly, we note that the terms *object-based* and *surface-based* attention have been used to refer to a variety of different effects in different paradigms. We think the mechanisms underlying the object-based effects used in the transparent-motion stimuli are likely different, at least in part, from that found in other paradigms. For instance, one (now classic) paradigm uses spatially separated stimuli and hence admits spatial selection based on the outline of the object in question (e.g., Chen, 1998; Egly et al., 1994; Lavie & Driver, 1996; Macquistan, 1997). Other studies have used superimposed objects but plausibly allow for object-class template mechanisms (Cohen & Tong, 2015). Further work is needed to determine the overlap in neuronal mechanisms that support these various attentional phenomena.

5. Conclusions

Our findings provide the first evidence of the role of the N1 component in object-based attention under conditions that rule out the feature-based explanations that have been shown to be admitted by previous studies (Stoner & Blanc, 2010). Together with previous research, these findings reinforce the importance of the N1 component in object-based selective processing as well as other types of attentional phenomena. Both our behavioral and EEG results provide evidence of an attentional processing advantage that spreads from one feature to another based on the spatiotemporal continuity of local texture elements (dots in the case of dot fields) rather than relying on higher-order mechanisms. Although we have speculated that this fine-grained spatial selection may involve area V1, future research is needed to identify the mechanisms underlying the behavioral and neuronal effects documented in this study.

Funding

This work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK BIDEB 2221 program) and the Turkish Academy of Sciences (TUBA-GEBIP Award). The author Esra Nur Catak was also supported by the National Scholarship Program for Ph.D. Students (TUBITAK BIDEB 2211-E program). The author Mert Özkan was also supported by National Science Foundation (NSF EPSCoR #1632738 grant).

Open practices

The study in this article earned Open Materials and Open Data badges for transparent practices. The data, materials and code for the study are available at https://osf.io/kpjgv/

CRediT authorship contribution statement

Esra N. Catak: Methodology; Formal analysis; Investigation; Data curation; Writing-original draft; Visualization. **Mert Ozkan:** Methodology; Software; Formal analysis; Investigation. **Hulusi Kafaligonul:** Conceptualization; Formal analysis; Writing-review & editing; Visualization; Supervision; Project administration; Funding acquisition. **Gene R. Stoner:** Conceptualization; Methodology; Software; Writing-review & editing; Supervision; Funding acquisition.

Declaration of competing interest

None.

Acknowledgements

We would like to thank Sibel Akyuz and Utku Kaya for suggestions on the data analyses. We are also grateful to Jude F. Mitchell for providing insightful and helpful comments on an earlier version of the manuscript.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cortex.2022.02.013.

REFERENCES

Akyuz, S., Pavan, A., Kaya, U., & Kafaligonul, H. (2020). Short- and long-term forms of neural adaptation: An ERP investigation of dynamic motion aftereffects. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior, 125*, 122–134.

Allen, P. J., Polizzi, G., Krakow, K., Fish, D. R., & Lemieux, L. (1998). Identification of EEG events in the MR scanner: The problem of pulse artifact and a method for its subtraction. *Neuroimage, 8*, 229–239.

Blaser, E., Pylyshyn, Z. W., & Holcombe, A. O. (2000). Tracking an object through feature space. *Nature, 408*(6809), 196–199.

Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Chen, Z. (1998). Switching attention within and between objects: The role of subjective organization. *Canadian Journal of Experimental Psychology, 52*, 7–16.

Ciaramitaro, V. M., Mitchell, J. F., Stoner, G. R., Reynolds, J. H., & Boynton, G. M. (2011). Object-based attention to one of two superimposed surfaces alters responses in human early visual cortex. *Journal of Neurophysiology, 105*(3), 1258–1265.
Clark, V. P., Fan, S., & Hillyard, S. A. (1995). Identification of early visual evoked potential generators by retinotopic and topographic analyses. *Human Brain Mapping, 2*, 170–187.
Cohen, E. H., & Tong, F. (2015). Neural mechanisms of object-based attention. *Cerebral Cortex, 25*(4), 1080–1092.
Di Russo, F., Martinez, A., & Hillyard, S. A. (2003). Source analysis of event-related cortical activity during visuo-spatial attention. *Cerebral Cortex, 13*(5), 486–499.
Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113*, 501–517.
Duncan, J. (1996). Cooperating brain systems in selective perception and action. In T. Inui, & J. L. McClelland (Eds.), *Attention and performance. Attention and performance 16: Information integration in perception and communication* (pp. 549–578). The MIT Press.
Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General, 123*(2), 161–167.
Ernst, Z. R., Boynton, G. M., & Jazayeri, M. (2013). The spread of attention across features of a surface. *Journal of Neurophysiology, 110*(10), 2426–2439.
Fallah, M., Stoner, G. R., & Reynolds, J. H. (2007). Stimulus-specific competitive selection in macaque extrastriate visual area V4. *Proceedings of the National Academy of Sciences, 104*(10), 4165–4169.
Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology, 48*, 1711–1725.
Hubel, D. H., & Wiesel, T. N. (1974). Sequence regularity and geometry of orientation columns in the monkey striate cortex. *Journal of Comparative Neurology, 158*(3), 267–293.
Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43*(3), 171–216.
Ives, H. E. (1912). XII. Studies in the photometry of lights of different colours. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 24*(139), 149–188.
Kafaligonul, H., Albright, T. D., & Stoner, G. R. (2018). Auditory modulation of spiking activity and local field potentials in area MT does not appear to underlie an audiovisual temporal illusion. *Journal of Neurophysiology, 120*, 1340–1355.
Kaya, U., & Kafaligonul, H. (2019). Cortical processes underlying the effects of static sound timing on perceived visual speed. *Neuroimage, 199*, 194–205.
Khoe, W., Mitchell, J. F., Reynolds, J. H., & Hillyard, S. A. (2005). Exogenous attentional selection of transparent superimposed surfaces modulates early event-related potentials. *Vision Research, 45*(24), 3004–3014.
Khoe, W., Mitchell, J. F., Reynolds, J. H., & Hillyard, S. A. (2008). ERP evidence that surface-based attention biases interocular competition during rivalry. *Journal of Vision, 8*(3), 18.
Krekelberg, B., & Albright, T. D. (2005). Motion mechanisms in macaque MT. *Journal of Neurophysiology, 93*(5), 2908–2921.
Lavie, N., & Driver, J. (1996). On the spatial extent of attention in object-based visual selection. *Perception & Psychophysics, 58*, 1238–1251.
Lee, J., & Maunsell, J. H. (2009). A normalization model of attentional modulation of single unit responses. *Plos One, 4*(2), Article e4651.
Lopez, M., Rodriguez, V., & Valdes-Sosa, M. (2004). Two-object attentional interference depends on attentional set. *International Journal of Psychophysiology, 53*(2), 127–134.
Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
Macquistan, A. (1997). Object based allocation of visual attention in response to exogenous, but not endogenous, spatial precues. *Psychonomic Bulletin & Review, 4*, 512–515.
Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190.
Mitchell, J. F., Stoner, G. R., Fallah, M., & Reynolds, J. H. (2003). Attentional selection of superimposed surfaces cannot be explained by modulation of the gain of color channels. *Vision Research, 43*(12), 1323–1328.
Mitchell, J. F., Stoner, G. R., & Reynolds, J. H. (2004). Object-based attention determines dominance in binocular rivalry. *Nature, 429*(6990), 410–413.
Müller, R., Göpfert, E., Breuer, D., & Greenlee, M. W. (1998). Motion VEPs with simultaneous measurement of perceived velocity. *Documenta Ophthalmologica, 97*(2), 121–134.
O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature, 401*, 584–587.
Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software or advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*, 156869.
Pelli, D. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.
Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology, 72*, 184–187.
Pinilla, T., Cobo, A., Torres, K., & Valdes-Sosa, M. (2001). Attentional shifts between surfaces: Effects on detection and early brain potentials. *Vision Research, 41*(13), 1619–1630.
Qian, N., & Andersen, R. A. (1994). Transparent motion perception as detection of unbalanced motion signals. II. Physiology. *Journal of Neuroscience, 14*(12), 7367–7380.
Recanzone, G. H., & Wurtz, R. H. (2000). Effects of attention on MT and MST neuronal activity during pursuit initiation. *Journal of Neurophysiology, 83*(2), 777–790.
Recanzone, G. H., Wurtz, R. H., & Schwarz, U. (1997). Responses of MT and MST neurons to one and two moving objects in the receptive field. *Journal of neurophysiology, 78*(6), 2904–2915.
Reynolds, J. H., Alborzian, S., & Stoner, G. R. (2003). Exogenously cued attention triggers competitive selection of surfaces. *Vision Research, 43*(1), 59–66.
Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience, 19*(5), 1736–1753.
Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron, 61*(2), 168–185.
Reynolds, J. H., Pasternak, T., & Desimone, R. (2000). Attention increases sensitivity of V4 neurons. *Neuron, 26*(3), 703–714.
Rodríguez, V., & Valdés-Sosa, M. (2006). Sensory suppression during shifts of attention between surfaces in transparent motion. *Brain Research, 1072*(1), 110–118.
Rodriguez, V., Valdes-Sosa, M., & Freiwald, W. (2002). Dividing attention between form and motion during transparent surface perception. *Cognitive Brain Research, 13*(2), 187–193.
Ruff, D. A., & Cohen, M. R. (2017). A normalization model suggests that attention changes the weighting of inputs between visual areas. *Proceedings of the National Academy of Sciences, 114*(20), E4085–E4094.
Schoenfeld, M. A., Hopf, J. M., Merkel, C., Heinze, H. J., & Hillyard, S. A. (2014). Object-based attention involves the sequential activation of feature-specific cortical modules. *Nature Neuroscience, 17*(4), 619–624.

Schoenfeld, M. A., Tempelmann, C., Martinez, A., Hopf, J. M., Sattler, C., Heinze, H. J., et al. (2003). Dynamics of feature binding during object-selective attention. *Proceedings of the National Academy of Sciences, 100*(20), 11806–11811.

Snowden, R. J., Treue, S., Erickson, R. G., & Andersen, R. A. (1991). The response of area MT and V1 neurons to transparent motion. *Journal of Neuroscience, 11*(9), 2768–2785.

Stoner, G. R. (2010). *Area V1 "point-set" as the unit of "object-based" selection*. Program No. 172.13. 2018 Neuroscience Meeting Planner. San Diego, CA: Society for Neuroscience. Online.

Stoner, G. R., & Blanc, G. (2010). Exploring the mechanisms underlying surface-based stimulus selection. *Vision Research, 50*(2), 229–241.

Sundberg, K. A., Mitchell, J. F., Gawne, T. J., & Reynolds, J. H. (2012). Attention influences single unit and local field potential response latencies in visual cortical area V4. *Journal of Neuroscience, 32*, 16040–16050.

Thaler, L., Schütz, A. C., Goodale, M. A., & Gegenfurtner, K. R. (2013). What is the best fixation target? The effect of target shape on stability of fixational eye movements. *Vision Research, 76*, 31–42.

Tootell, R., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., et al. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *Journal of Neuroscience, 15*(4), 3215–3230.

Ts'o, D. Y., Zarella, M., & Burkitt, G. (2009). Whither the hypercolumn? *The Journal of Physiology, 587*(12), 2791–2805.

Valdes-Sosa, M., Bobes, M. A., Rodriguez, V., Acosta, Y., Perez, A., Iglesias, J., et al. (2004). The influence of scene organization on attention: Psychophysics and electrophysiology. In N. Kanwisher, & J. Duncan (Eds.), *Attention and performance XX: Functional neuroimaging of visual cognition* (pp. 321–344). Oxford, UK: Oxford University Press.

Valdes-Sosa, M., Bobes, M. A., Rodriguez, V., & Pinilla, T. (1998). Switching attention without shifting the spotlight: Object-based attentional modulation of brain potentials. *Journal of Cognitive Neuroscience, 10*(1), 137–151.

Valdes-Sosa, M., Cobo, A., & Pinilla, T. (1998). Transparent motion and object-based attention. *Cognition, 66*, B13–B23.

Valdes-Sosa, M., Cobo, A., & Pinilla, T. (2000). Attention to object files defined by transparent motion. *Journal of Experimental Psychology: Human Perception and Performance, 26*(2), 488.

Verhoef, B. E., & Maunsell, J. H. (2016). Attention operates uniformly throughout the classical receptive field and the surround. *Elife, 5*, Article e17256.

Wanning, A., Rodriguez, V., & Freiwald, W. A. (2007). Attention to surfaces modulates motion processing in extrastriate area MT. *Neuron, 54*(4), 639–651.

Watson, J. D., Myers, R., Frackowiak, R. S., Hajnal, J. V., Woods, R. P., Mazziotta, J. C., & Zeki, S. (1993). Area V5 of the human brain: Evidence from a combined study using positron emission tomography and magnetic resonance imaging. *Cerebral cortex, 3*(2), 79–94.

Williford, T., & Maunsell, J. H. (2006). Effects of spatial attention on contrast response functions in macaque area V4. *Journal of Neurophysiology, 96*(1), 40–54.

World Medical Association. (2013). Declaration of Helsinki: Ethical principles for medical research involving human subjects. *Journal of the American Medical Association, 310*(20), 2191–2194.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology. Human Perception and Performance, 10*(5), 601.

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology. Human Perception and Performance, 16*(1), 121.