Behavioral/Cognitive

Task-Irrelevant Visual Forms Facilitate Covert and Overt Spatial Selection

Amarender R. Bogadhi,[1,2] Antimo Buonocore,[1,2] and Ziad M. Hafed[1,2]
[1]Hertie Institute for Clinical Brain Research, University of Tuebingen, Tuebingen, Germany, 72076, and [2]Werner Reichardt Centre for Integrative Neuroscience, University of Tuebingen, Tuebingen, Germany, 72076

Covert and overt spatial selection behaviors are guided by both visual saliency maps derived from early visual features as well as priority maps reflecting high-level cognitive factors. However, whether mid-level perceptual processes associated with visual form recognition contribute to covert and overt spatial selection behaviors remains unclear. We hypothesized that if peripheral visual forms contribute to spatial selection behaviors, then they should do so even when the visual forms are task-irrelevant. We tested this hypothesis in male and female human subjects as well as in male macaque monkeys performing a visual detection task. In this task, subjects reported the detection of a suprathreshold target spot presented on top of one of two peripheral images, and they did so with either a speeded manual button press (humans) or a speeded saccadic eye movement response (humans and monkeys). Crucially, the two images, one with a visual form and the other with a partially phase-scrambled visual form, were completely irrelevant to the task. In both manual (covert) and oculomotor (overt) response modalities, and in both humans and monkeys, response times were faster when the target was congruent with a visual form than when it was incongruent. Importantly, incongruent targets were associated with almost all errors, suggesting that forms automatically captured selection behaviors. These findings demonstrate that mid-level perceptual processes associated with visual form recognition contribute to covert and overt spatial selection. This indicates that neural circuits associated with target selection, such as the superior colliculus, may have privileged access to visual form information.

Key words: **covert attention; object recognition; overt attention; priority maps; saccades; visual form recognition**

Significance Statement

Spatial selection of visual information either with (overt) or without (covert) foveating eye movements is critical to primate behavior. However, it is still not clear whether spatial maps in sensorimotor regions known to guide overt and covert spatial selection are influenced by peripheral visual forms. We probed the ability of humans and monkeys to perform overt and covert target selection in the presence of spatially congruent or incongruent visual forms. Even when completely task-irrelevant, images of visual objects had a dramatic effect on target selection, acting much like spatial cues used in spatial attention tasks. Our results demonstrate that traditional brain circuits for orienting behaviors, such as the superior colliculus, likely have privileged access to visual object representations.

Introduction

Spatial selection of stimuli in a cluttered visual scene is central to visual behaviors in primates, and it could occur either overtly (with orienting eye movements) or covertly (without such eye movements). The mechanisms underlying both overt and covert spatial selection behaviors rely on spatial maps in sensorimotor regions that are functionally organized into visual saliency maps, primarily derived from low-level visual processes, and priority maps, representing high-level cognitive processes (Fecteau and Munoz, 2006; Veale et al., 2017; Bisley and Mirpour, 2019). Indeed, classic visual saliency map models are computed from early visual features, such as orientation, motion, and color (Itti and Koch, 2000), whereas priority maps are based on cognitive factors, such as behavioral relevance, expectation, and reward (Awh et al., 2012; Chelazzi et al., 2014; Sprague et al., 2018). Accordingly, visual saliency maps and priority maps are believed to be represented in the neuronal activity of visual, sensorimotor, and associative brain regions, such as primary visual cortex (V1), superior colliculus (SC), and regions of the parietal, and prefrontal cortices (Gottlieb et al., 1998; Bisley and Goldberg,

Received June 25, 2020; revised Sep. 8, 2020; accepted Oct. 7, 2020.

Author contributions: A.R.B. and Z.M.H. designed research; A.R.B. and A.B. performed research; A.R.B. analyzed data; A.R.B. wrote the first draft of the paper; A.R.B., A.B., and Z.M.H. edited the paper; A.R.B. and Z.M.H. wrote the paper.

The authors declare no competing financial interests.

This work was supported by Werner Reichardt Center for Integrative Neuroscience, Deutsche Forschungsgemeinschaft EXC 307 excellence cluster, Hertie Institute for Clinical Brain Research, and Deutsche Forschungsgemeinschaft Project BO5681/1-1.

Correspondence should be addressed to Amarender R. Bogadhi at bogadhi.amar@gmail.com.

https://doi.org/10.1523/JNEUROSCI.1593-20.2020

Copyright © 2020 the authors

2003; Ignashchenkova et al., 2004; White et al., 2017; Sapountzis et al., 2018; Yan et al., 2018; Chen et al., 2020).

The organization of spatial maps based on a dichotomy of early visual features, on the one hand, and high-level cognitive factors, on the other, ignores whether mid-level perceptual processes related to visual form recognition are represented in these maps. This is inconsistent with multiple lines of evidence suggesting a potential link between visual form recognition and spatial selection behaviors. First, recent studies in monkeys identified a novel attention-related region in the temporal cortex (Bogadhi et al., 2019a; Stemmann and Freiwald, 2019). Importantly, neurons in this attention-related region were selective for peripheral visual forms (Bogadhi et al., 2019b), suggesting a functional link between covert spatial selection and peripheral visual form recognition. Second, studies modeling fixation patterns in free-viewing of natural images show that visual objects predict fixation patterns better than saliency maps based on early visual features (Einhäuser et al., 2008; Yanulevskaya et al., 2013; Kümmerer et al., 2014), indicating an influence of visual forms on overt behaviors. Third, behavioral studies in humans show a rapid detection of faces and animals in peripheral images for saccadic eye movements and attentional capture, suggesting a rapid processing of animate visual forms for overt selection (Kirchner and Thorpe, 2006; Bindemann et al., 2007; Crouzet et al., 2010; Drewes et al., 2011; Devue et al., 2012). However, such rapid detection could be explained by low-level image features or unnatural statistics of the image databases (Honey et al., 2008; Wichmann et al., 2010; Crouzet and Thorpe, 2011; Zhu et al., 2013). Importantly, these studies used visual form images as saccade targets, which were always relevant to the task performance. Hence, it remains unclear, from studies using goal-directed and free-viewing paradigms, whether peripheral visual forms contribute to spatial selection when they are rendered irrelevant to the task and equated with nonform images for low-level visual features. We hypothesized that, if peripheral visual forms contribute to spatial selection, then they should do so in both covert and overt behaviors, even when the visual forms are task-irrelevant and equated for low-level visual features.

We investigated the contribution of peripheral visual forms to covert and overt spatial selection using a visual detection task pitting visual form images against 50% phase-scrambled images. Most importantly, all images were irrelevant to the task and matched for low-level image properties. In the covert (humans) and overt (humans and monkeys) tasks, subjects reported the detection of a suprathreshold target with a manual or saccadic eye movement response, respectively. We found that response times were significantly faster when the target was congruent with a visual form image in both covert and overt selection tasks, and in both humans and monkeys. Crucially, almost all response errors were captured by visual forms incongruent with targets. Interestingly, during covert selection, microsaccades following image onsets were biased toward visual forms. These findings demonstrate that peripheral visual forms, even when task-irrelevant, contribute to overt and covert spatial selection and perhaps act as spatial cues for orienting movements (Posner, 1980; Tian et al., 2016).

Materials and Methods

Subjects and ethics approvals

Eleven human subjects (3 males and 8 females; mean age ± SD = 27.3 ± 3.9 years) naive to the purpose of the study and 3 male rhesus monkeys (*Macaca mulatta*; Monkeys A, F, and M) aged 10, 11, and 10 years, respectively, participated in this study. All human subjects provided written informed consent in accordance with the Declaration of Helsinki. Ethics committees at the Medical Faculty of Tuebingen University reviewed and approved protocols for the human experiments. Monkey experiments were approved by regional governmental offices in Tuebingen.

Experimental setups

Human subjects were seated in a dark room at a viewing distance of 57 cm from a CRT monitor with a resolution of 1400 × 1050 pixels (34.13° × 25.93°). Stimulus display on the monitor was controlled by a 2010 Mac Pro (Apple) running MATLAB (The Mathworks) with the Psychophysics Toolbox extensions (Brainard, 1997). Eye position signals and manual responses were acquired using an EyeLink 1000 infrared eye-tracking system (SR Research) and Viewpixx button box (VPixx Technologies), respectively.

Monkeys were seated and head-fixed in a primate chair (Crist Instrument) inside a darkened booth at a viewing distance of 72.2 cm from a CRT monitor with a 1024 × 768 resolution (30.96° × 23.47°). For experiments in Monkeys A and M, stimulus display on the monitor was controlled using a modified version of PLDAPS with Datapixx and Psychophysics Toolbox extensions on MATLAB (The MathWorks) running on an Ubuntu operating system (Eastman and Huk, 2012). For experiments in Monkey F, stimulus display was controlled using a LabVIEW system (National Instruments) handshaking with a 2010 Mac Pro (Apple) running MATLAB (The MathWorks) with the Psychophysics Toolbox extensions (for details, see Chen and Hafed, 2013; Tian et al., 2016). Eye position signals in Monkeys A and M were measured using a surgically implanted scleral search coil; eye position signals in Monkey F were measured using an EyeLink 1000 infrared eye-tracking system (SR Research). Surgical procedures for implanting head-holders and scleral coils were described in a previous study (Skinner et al., 2019).

Experimental design

Covert selection task (human subjects). Subjects started each trial by fixating a central spot of 0.1° radius (97.6 cd/m^2) displayed on a gray background (43.84 cd/m^2). Eye position signals were monitored to enforce fixation within a fixation window of 2° radius. Following a 500–1000 ms randomized fixation duration, an intact visual form image (4.88° × 4.88°) and its corresponding 50% scrambled form image (see Image normalization) were displayed symmetrically on either side of fixation along the horizontal meridian and centered at 8° eccentricity. After a fixed delay of 100, 200, or 300 ms following image onset, a suprathreshold target (black disk; radius = 0.2°) was displayed at the center of one of the two images. Subjects were instructed to report the spatial location of the target with a left or right button press; and most importantly, they were informed that both images were irrelevant to the detection task. We refer to trials in which the target was presented on top of the visual form image as target congruent trials and trials in which the target was presented on top of the 50% phase-scrambled image as target incongruent trials. All three delay conditions (100, 200, or 300 ms) were randomized across trials. In addition, catch trials with no target were also included in the 100 and 300 ms delay conditions on 25% of trials, and subjects were instructed to withhold their responses in these trials. Data in each subject ($n = 11$) were collected in a single session.

Overt selection task (humans and monkeys). The trial structure in the overt task was the same as in the covert task with one difference: in the overt task, subjects reported the detection of the target with a saccade to the target location rather than a button press. In addition, at the same time as target onset, the fixation point was extinguished and presented on top of the target to aid the subjects with a speeded saccade response to the target location. We suggest that this was particularly helpful in instructing monkeys that would otherwise require additional training to generate target-directed saccades. However, it is important to note that fixation point disappearance was at the center of the display and was completely uninformative of the newly appearing target location. Crucially, the identical visual event happened at the same location in both the target congruent and target incongruent trials that we compared throughout this study.

We also included a single-image condition on 40% of trials. In this case, only one image, either a visual form image or a 50% scrambled

form image, was presented simultaneously with a target on top of it in one of the four diagonal locations at 8° eccentricity. The single-image condition with diagonal locations was used to control for any spatial biases in eye movements from repetitive target presentations at the same spatial locations. We collected data in 10 of the 11 human subjects in a single session each and in 3 monkeys across 17 sessions. In monkeys, we used the same visual images (now sized 5.14° × 5.14°) as in the human experiments, with a target (black disk) of radius 0.3°–0.45° across monkeys. The background gray luminance for the monkey experiments was 27.21–37.1 cd/m^2. The humans performed the overt selection task before performing the covert selection task.

The randomized fixation duration before image presentation for the overt task in humans was the same as in the covert task (500–1000 ms). However, the duration was slightly different across monkeys. In Monkeys A and M, the duration was 100–700 ms compared with 300–900 ms in Monkey F. This was because of the slightly different experimental setup in which data from Monkey F were collected. Nonetheless, to keep the timing of image presentations comparable across monkeys in all analyses, we excluded trials in Monkeys A and M having fixation durations <300 ms. This trial exclusion was blind to whether the trials were target congruent or target incongruent trials. Most importantly, we have repeated the analyses on all trials (i.e., with no exclusion), and the results were unaltered (see Results).

Image normalization

Forty images with visual forms and their corresponding 50% scrambled form images were used in both human and monkey experiments. Visual form images were obtained from previous electrophysiology studies of the inferotemporal cortex (Tsao et al., 2006; Bogadhi et al., 2019b) and were sampled from four different categories, including human faces, fruits, hands, and inanimate objects with 10 examples in each category (see Fig. 1*c*).

All images were equated iteratively for luminance distributions (mean = background luminance) and Fourier spectra using the SHINE tool box (Willenbockel et al., 2010). Briefly, all 40 visual form images were resized to the appropriate dimensions, and the mean gray level of each image was equated to the background level (*lumMatch in SHINE*). The resultant 40 images were iteratively ($n = 20$) matched for the histogram of gray levels (*histMatch in SHINE*) and the Fourier spectra (*specMatch in SHINE*), before generating their corresponding 50% phase-scrambled images by randomizing half of the phase matrix and keeping the amplitude matrix constant. Finally, all of the visual form images and their corresponding phase-scrambled images were iteratively ($n = 20$) matched, once again, for the histogram of gray levels and the Fourier spectra to yield the final visual form images (see Fig. 1*c*, top) and phase-scrambled images (see Fig. 1*c*, bottom) used in this study.

Suprathreshold target detection

We hypothesized that visual form contribution to spatial selection behaviors, in covert and overt tasks, should be evident as a facilitation of response times between target congruent and target incongruent conditions. Hence, it was important that the differences in response times between target congruent and target incongruent conditions could not be attributed to difficulty in the visual detection of the target across conditions. For this reason, we chose a high-contrast ("black" color) and sufficiently large target (0.2° radius disk). We also confirmed that detection performance during the most difficult condition of our covert task (the 100 ms delay condition; see Results) was suprathreshold in both target congruent (% correct performance = 99.48 ± 0.97% SD) and target incongruent (% correct performance = 99.06 ± 0.94% SD) trials, with no significant difference between the two conditions (Wilcoxon signed-rank test, $p = 0.43$).

Statistical analyses

Response times and proportions of errors. We measured visual form contribution to spatial selection behaviors on correct and error trials separately. Correct trials were defined as the trials in which the first response of the subject correctly matched the target location. Error trials were those in which the subjects erroneously selected the image that had no target dot superimposed on it (i.e., they selected the image opposite to the target). On correct trials, we quantified the effect of visual forms on response time differences between target congruent and target incongruent conditions. On error trials, we quantified the effect of visual forms on the proportion of errors between target congruent and target incongruent conditions. Response time on a correct trial was calculated as the time of response onset relative to target onset. The proportion of

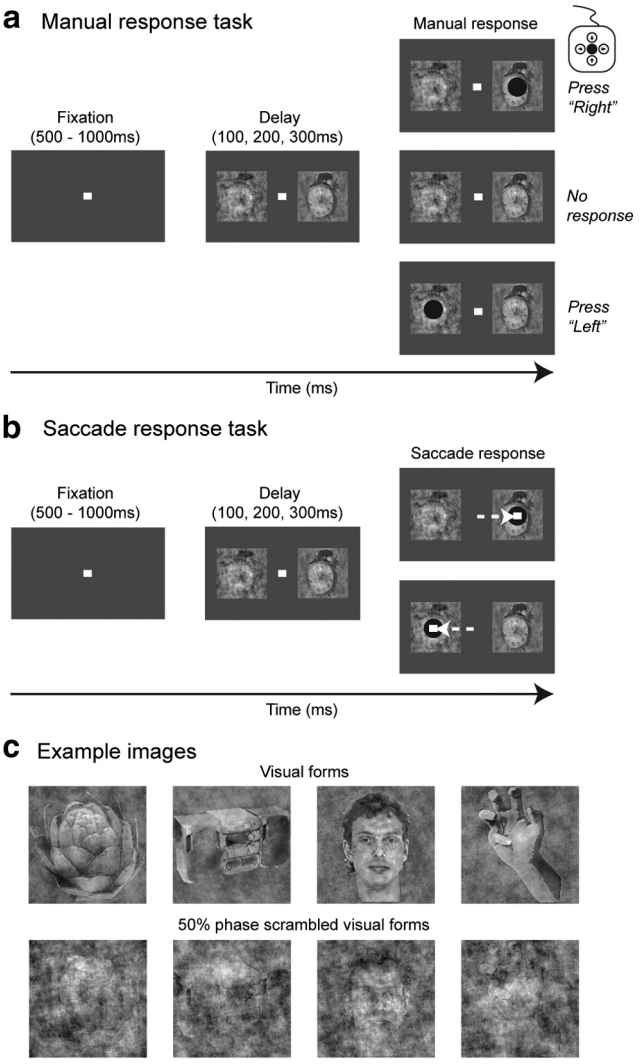

Figure 1. Covert (manual response) and overt (saccade response) selection tasks. ***a***, Trial epochs in the covert selection task. Each trial started with a fixation spot at the center of the screen. After 500-1000 ms of fixation, two images (one with an intact visual form and the other with the corresponding 50% phase-scrambled visual form) appeared on either side of fixation along the horizontal meridian and centered at 8° eccentricity. Following a fixed delay of 100, 200, or 300 ms, a suprathreshold target ("black disk") appeared on top of one of the images, and subjects were instructed to report the detection of the target with an appropriate button press. On the remaining "catch" trials (see Materials and Methods), no target was presented, and subjects were required to withhold their response. ***b***, Trial epochs in the overt selection task. The task consisted of instruction ("single image") and distractor ("two images") trials. The distractor trials shown here were similar to the covert selection task until the target presentation. After a fixed delay of 100, 200, or 300 ms from target onset, a suprathreshold target ("black disk") appeared on top of one of the images, and the subjects were instructed to detect the target and generate a saccade to its location. At the time of target appearance, the fixation spot at the center was extinguished and was presented on top of the target to instruct the subjects to generate a visually guided saccade. ***c***, Example images of intact visual forms (top row) and their corresponding 50% phase-scrambled visual forms (bottom row) from the four categories used in this study. A total of 40 visual form images, 10 from each category, were used after equating for low-level features (see Materials and Methods).

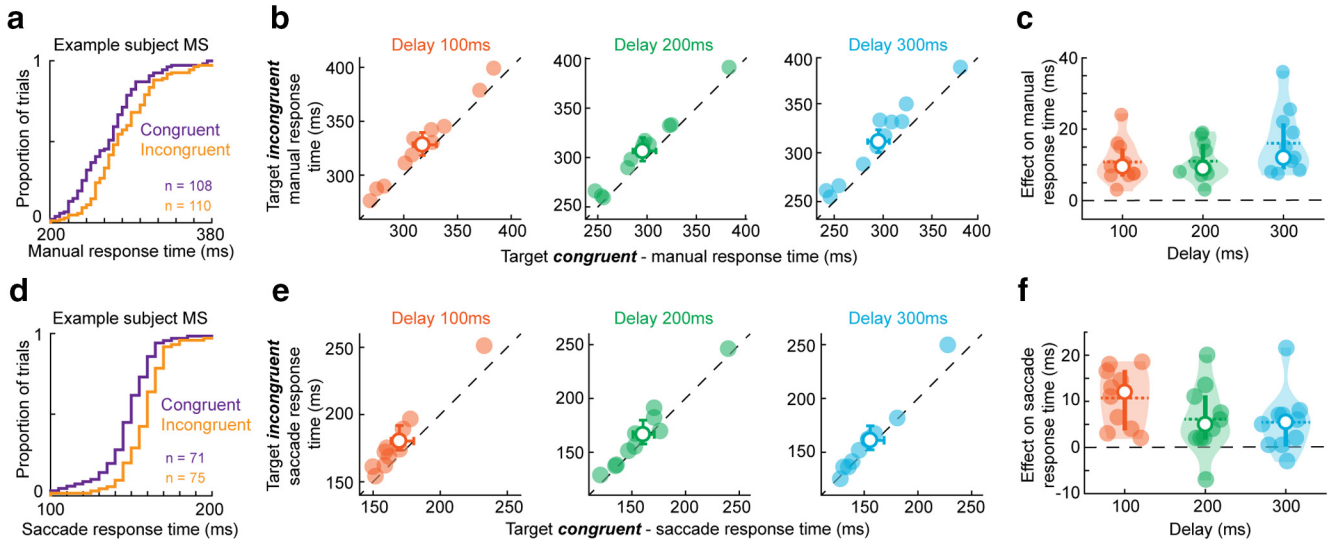

Figure 2. Visual form facilitation of response times in covert and overt selection tasks in humans. ***a***, Cumulative distribution of manual response times in an example subject for target congruent and target incongruent trials in the covert selection task during the fixed delay interval of 100 ms (see Materials and Methods). The number of samples in each cumulative distribution is shown in the panel. ***b***, Each panel represents a paired comparison of median response times between target congruent and target incongruent trials in the covert selection task across all subjects ($n = 11$; filled circles) for a fixed delay of 100 ms (red), 200 ms (green), or 300 ms (blue). Colored circle with errors bars in each panel represents mean ± SD across subjects. Dotted line indicates the line of unity slope. Data above the line indicated faster response times in the target congruent condition. ***c***, The facilitatory effect of visual forms on manual response times in the covert task was quantified in each subject as a difference in median response times between target incongruent and target congruent trials (same data as in ***b***). Each data point in the violin plots is from each subject. Colored circles and horizontal bars in each violin plot represent median and mean, respectively. Vertical bars represent SD. Same color conventions as in ***b***. ***d-f***, Same as in ***a-c***, but now for the saccade responses of humans in the overt selection task.

errors in each subject was calculated as the ratio of the number of error trials to the sum of correct and error trials.

Saccadic responses to targets in the overt task and microsaccades during fixation in both the covert and overt tasks were detected using a velocity and acceleration threshold followed by manual inspection (Krauzlis and Miles, 1996). Trials with microsaccades occurring between 100 ms before image onset and response onset were excluded in the analyses of response times and proportion of errors to control for the effects of microsaccades on stimulus onset activity in different brain regions (Chen et al., 2015).

For analysis of response times in the covert task in humans (e.g., see Fig. 2*a*), we included an average of 80.6 (SD = 26.4) and 77.9 (SD = 24.9) trials from target congruent and target incongruent conditions, respectively, for a given delay and subject. Similarly, in the overt task (see Fig. 2*b*), we included an average of 68.2 (SD = 9.8) and 68.6 (SD = 9.8) trials in target congruent and target incongruent conditions, respectively. For the response time analysis in monkeys, trial counts in target congruent and target incongruent conditions are shown in Figure 6. It should be noted here that there was a progressively lower trial count with increasing delay in the monkey data (see Fig. 6). This is primarily because of the exclusion of trials with microsaccades occurring between 100 ms before image onset and 100 ms after target onset. That is, the longer the delay period duration, the more likely it was for microsaccades to have occurred during our exclusion period. This progressive reduction in trial count was also more apparent in Monkeys A and M compared with Monkey F. We measured the microsaccade rate in all 3 monkeys after the image onset, from 150 to 300 ms, and confirmed that Monkeys A and M generated more microsaccades compared with Monkey F (microsaccades/s: mean ± SD = 0.79 ± 0.36 and 1.15 ± 0.47 for Monkeys A and M respectively, vs 0.44 ± 0.13 for Monkey F), explaining the greater loss of trials in Monkeys A and M. Nonetheless, we verified that trial count before exclusion was comparable across the three delay conditions in all monkeys (mean ± SD = 527.66 ± 15.13 in Monkey A, 637.5 ± 8.31 in Monkey F, 183.66 ± 13.53 in Monkey M). More importantly, we repeated our analyses without any microsaccade exclusion (i.e., with matched trial numbers across delays), and the results remained consistent with our main findings (see Results). For all paired comparisons in humans and monkeys, we used Wilcoxon signed-rank tests.

Microsaccades. We pooled microsaccades flagged during the covert task across subjects, and we separated them into congruent and incongruent microsaccades based on their directions relative to the visual form image or the scrambled image, respectively. Specifically, we found that the great majority of microsaccades were predominantly horizontal because of our image configuration in the task. We therefore grouped all microsaccades within ±32 degrees from horizontal into two groups depending on the direction of their horizontal component: either toward the visual form image (congruent microsaccades) or toward the scrambled image (incongruent microsaccades; see Fig. 4*a*, inset). The rate of congruent and incongruent microsaccades was constructed by counting the corresponding microsaccades in a 50 ms time window sliding in steps of 5 ms. The proportion of microsaccades congruent with visual forms was also calculated as the ratio of the number of congruent microsaccades to the total sum of congruent and incongruent microsaccades occurring within a given time bin. To test the statistical significance of congruent and incongruent microsaccade proportions, we used the binomial test in each of the 50 ms time windows. These analyses of microsaccade rate and direction congruence are standard analyses in the field of microsaccade research (Engbert and Kliegl, 2003; Pastukhov and Braun, 2010; Hafed, 2013; Pastukhov et al., 2013; Tian et al., 2016; Baumeler et al., 2020). Also, in the overt task, saccades to the images replaced poststimulus microsaccades; it was therefore not meaningful to analyze microsaccadic modulations in the overt task.

ANOVAs. To test the influence of visual form category on the manual response times in humans and saccade response times in humans and monkeys, we performed ANOVAs on the response time data with three factors: visual forms (intact/scrambled), delay (100, 200, 300 ms), and categories (human faces, fruits, hands, and inanimate objects). In the ANOVA of response times in humans, we included the subjects as a random effect factor.

Results

Both covert and overt selection behaviors are facilitated by task-irrelevant visual form images

We hypothesized that, if peripheral visual forms contribute to spatial selection behaviors in an automatic and bottom-up

manner, then response times associated with target detection should be influenced by the spatial congruency between target location and task-irrelevant visual form images. To test this, we first ran human subjects on two target detection tasks: one being covert and involving a manual response (Fig. 1*a*) and the other being overt and using a foveating eye movement response (Fig. 1*b*). In both tasks, the subjects had to report, as quickly as possible, the onset of a suprathreshold target stimulus that appeared at one of two possible locations centered on top of either a visual form image or its scrambled version (see Materials and Methods). The subjects were informed *a priori* that the images behind the possible two target locations were completely irrelevant to the task, and we identified correct trials as those in which the required response (one of two buttons corresponding to each target location or accurate saccade landing at the target location) was spatially accurate; this was the majority of trials (see Materials and Methods). In all cases, the target could appear after one of three possible delays after image onset (Fig. 1). We compared response times on correct trials when the target was congruent with the visual form image to response times when the target was incongruent with the visual form image in each delay condition.

A comparison of cumulative distributions of response times in an example subject, during the 100 ms delay condition, clearly shows that response times for congruent targets were faster compared with incongruent targets in both the covert (Fig. 2*a*) and overt (Fig. 2*d*) tasks. A paired comparison of median response times across all subjects further demonstrates that response times for congruent targets were significantly faster compared with incongruent targets, in all three delay conditions (100, 200, and 300 ms) tested, and in both covert (Fig. 2*b*; Wilcoxon signed-rank test, $p = 0.0009$ in 100 ms, $p = 0.0009$ in 200 ms, $p = 0.0009$ in 300 ms) and overt (Fig. 2*e*; Wilcoxon signed-rank test, $p = 0.0019$ in 100 ms, $p = 0.027$ in 200 ms, $p = 0.011$ in 300 ms) tasks. This facilitation of response times by the visual form was uniformly present across all delays in the covert task (Fig. 2*c*), and it was the strongest for the 100 ms delay condition in the overt task (Fig. 2*f*). These findings demonstrate that peripheral visual forms, even when task-irrelevant, bias spatial selection and facilitate target detection as early as 100 ms from image onset, in both covert and overt spatial selection behaviors.

Visual forms capture selection even when incongruent with task requirements

In its complementary form, a spatial selection bias by visual forms could also degrade performance and result in more error trials when visual forms are spatially incongruent with target locations. We tested for this by comparing the proportion of errors in target congruent trials with the proportion of errors in target incongruent trials, in both the covert (Fig. 3*a*) and overt (Fig. 3*b*) tasks. On average, our subjects made more errors when the targets were incongruent with visual forms compared with when they were congruent with visual forms, and this occurred in both the covert (Fig. 3*a*) and overt (Fig. 3*b*) versions of the task. A paired comparison across subjects revealed a consistent pattern of more errors for incongruent targets in the 200 ms (Wilcoxon signed-rank test; covert task, $p = 0.03$; overt task, $p = 0.01$) and 300 ms (Wilcoxon signed-rank test; covert task, $p = 0.09$; overt task, $p = 0.06$) delay conditions, in both the covert (Fig. 3*a*, middle, right) and overt (Fig. 3*b*, middle, right) tasks. Interestingly, this effect of visual forms on errors for incongruent targets was weaker and less consistent across subjects in the 100 ms delay condition (Wilcoxon signed-rank test; covert task, $p = 0.43$; overt task, $p = 0.22$), and in both covert (Fig. 3*a*, left) and overt (Fig. 3*b*, left) tasks. These findings provide a complementary demonstration that peripheral visual forms bias spatial selection and produce more errors for incongruent targets in a time-specific manner (Fig. 2).

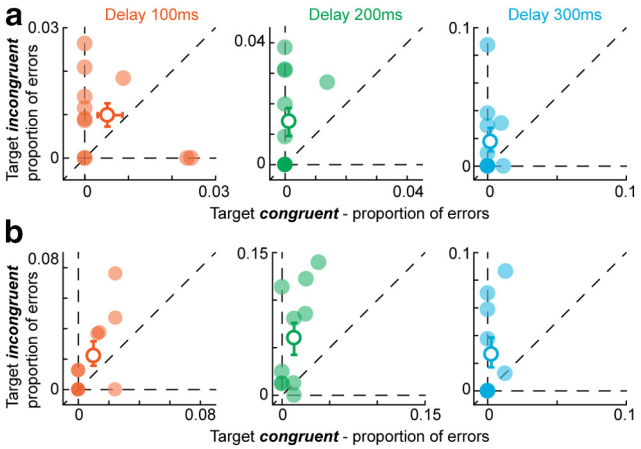

Figure 3. Visual form influence on errors in the covert and overt selection tasks in humans. ***a***, Each panel represents a paired comparison of the proportion of manual response errors between target congruent and target incongruent trials in the covert selection task across all subjects ($n = 11$; filled circles) for a fixed delay of 100 ms (red), 200 ms (green), or 300 ms (blue). Colored circle with errors bars in each panel represents mean ± SD across subjects. Dotted line indicates line of unity slope. Data above the line indicate more errors in the target incongruent condition. ***b***, Same as in ***a***, but now for the saccade response errors in the overt selection task.

Microsaccades during fixation reflect capture by peripheral, task-irrelevant visual forms

The behavioral effects of peripheral visual forms, particularly on response times (Fig. 2), in both covert and overt tasks resemble the well-known effects of spatial cues on behavioral performance in attention tasks (Posner, 1980). Since microsaccades provide a sensitive assay of effects related to attention (Hafed and Clark, 2002; Engbert and Kliegl, 2003), we therefore tested whether peripheral visual forms also bias microsaccades before target presentation. We analyzed the incidence of microsaccades before target presentation either toward (congruent) or opposite (incongruent) the suddenly appearing visual form image in the covert version of the task (see Materials and Methods). We first computed a microsaccade rate independently for movements that were either congruent with the visual form image or incongruent with it (Fig. 4*a*). Immediately after image onset, microsaccade rate for both congruent and incongruent movements decreased reflexively, consistent with previous reports of microsaccadic inhibition (Engbert and Kliegl, 2003; Rolfs et al., 2008; Tian et al., 2016; Buonocore et al., 2017). However, subsequent microsaccades, which likely benefit from frontal cortical drive (Peel et al., 2016), occurred earlier if they were congruent with a visual form than if they were incongruent (Fig. 4*a*). Importantly, this meant that the proportion of microsaccades in the congruent direction was higher than in the incongruent direction during the interval following inhibition, suggesting a spatial direction bias toward visual form images. This spatial bias was statistically different from chance in each of the 50 ms time bins from 122.5 to 142.5 ms (Fig. 4*b*; binomial test, $p < 0.05$). These findings demonstrate that peripheral visual forms, even when task irrelevant, bias microsaccades and effectively act as spatial cues for selection behaviors.

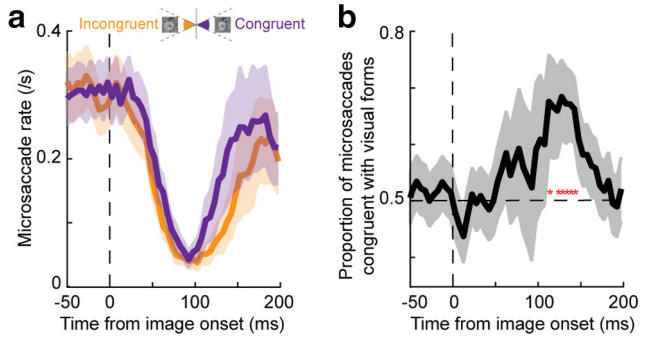

Figure 4. Visual form influence on human microsaccade directions in the covert selection task. ***a***, Mean microsaccade rate aligned to the time of image onset (vertical dotted line). Each curve indicates the rate of microsaccades that were in the same (congruent; orange) or opposite (incongruent; purple) directions from the visual form image (see Materials and Methods). Inset, Diagram of visual angle (32°) subtended by the images for calculating congruent and incongruent microsaccade directions. Error bars indicate bootstrapped 68.2% CIs. ***b***, Microsaccades in the congruent and incongruent directions (same data as in ***a***) were used to show the proportion of microsaccades biased in the direction of visual forms 100 ms after image onset (see Materials and Methods). Horizontal dotted line indicates chance level. Error bars indicate bootstrapped 68.2% CIs. Red colored symbols (*) indicate significance in the corresponding time bin using the binomial test ($p < 0.05$; see Materials and Methods).

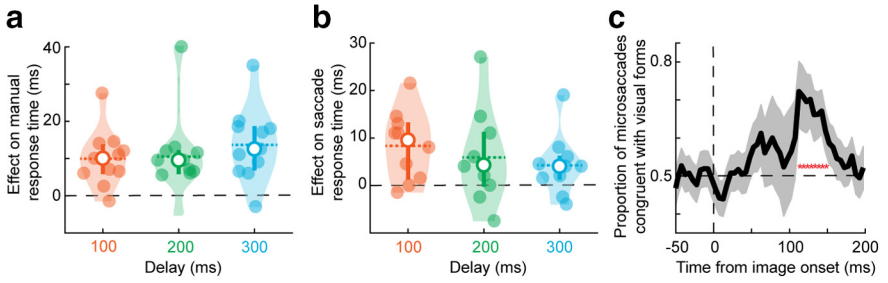

Figure 5. Nonface visual form facilitation of response times in covert and overt selection tasks in humans. ***a***, A facilitatory effect of nonface visual forms on manual response times in the covert selection task for the three delay conditions. Same conventions as in Figure 2*c*. ***b***, A facilitatory effect of nonface visual forms on saccade response times in the overt selection task for the three delay conditions. Same conventions as in Figure 2*f*. ***c***, Proportion of microsaccades biased in the direction of nonface visual forms after image onset, but before target onset (see Materials and Methods). Error bars indicate bootstrapped 68.2% CIs. ***c***, Red asterisk indicates significance in the corresponding time bin using the binomial test ($p < 0.05$; see Materials and Methods). Same conventions as in Figure 4*b*.

Nonface visual forms still influence response times and bias microsaccades

The known influence of face stimuli on saccadic eye movements (Bindemann et al., 2007; Xu-Wilson et al., 2009; Morand et al., 2010; Devue et al., 2012; Boucart et al., 2016; Kauffmann et al., 2019; Buonocore et al., 2020) raises a potential question on our results so far: namely, whether our findings of visual form effects on response times, microsaccade biases, and target selection errors are largely restricted to trials with face images. To test this, we excluded trials with face images and reanalyzed all of our data with only nonface images in both covert and overt tasks. We found that response times were faster for targets congruent with nonface visual forms compared with incongruent targets in both covert (Fig. 5*a*; Wilcoxon signed-rank test, $p = 0.002$ in 100 ms, $p = 0.002$ in 200 ms, $p = 0.002$ in 300 ms) and overt tasks (Fig. 5*b*; Wilcoxon signed-rank test, $p = 0.01$ in 100 ms, $p = 0.09$ in 200 ms, $p = 0.05$ in 300 ms). Importantly, we also observed a spatial direction bias in microsaccades before the target presentation toward nonface visual forms in each of the 50 ms bins from 112.5 to 147.5 ms (Fig. 5*c*; binomial test, $p < 0.05$). These results show that nonface visual forms strongly influence spatial selection to facilitate target detection in both covert and overt behaviors.

Additionally, we evaluated the complementary effect of nonface visual forms on errors when they were incongruent with the targets. Surprisingly, we found the effect of incongruent visual forms on errors to be weak and inconsistent across three delay conditions in both covert (Wilcoxon signed-rank test, $p = 0.74$ in 100 ms, $p = 0.21$ in 200 ms, $p = 0.12$ in 300 ms) and overt tasks (Wilcoxon signed-rank test, $p = 0.68$ in 100 ms, $p = 0.007$ in 200 ms, $p = 0.12$ in 300 ms). These findings suggest that nonface visual forms bias spatial selection only to an extent where it can facilitate spatially congruent target detection but not necessarily degrade spatially incongruent target detection.

We also tested whether there might be an influence of object category on the visual form effects on manual and saccade response times in humans (Fig. 2). That is, it could be possible that certain visual form categories (e.g., inanimate objects) are less ecologically relevant than other visual form categories (e.g., faces or fruits), and therefore have smaller effects on response times in our tasks. To investigate this, we ran ANOVAs with visual form, delay, and object categories as factors (see Materials and Methods). The results from our ANOVAs showed a main effect of visual form factor on response times in both manual ($F_{(1,4968)} = 67.42$, $p < 0.0001$, ANOVA) and saccade ($F_{(1,3866)} = 33.09$, $p = 0.0003$, ANOVA) response tasks, consistent with our main findings (Fig. 2). In addition, we also observed a significant interaction effect of visual form and category factors on saccade response times ($F_{(3,3866)} = 3.54$, $p = 0.027$, ANOVA), but not on manual response times ($F_{(3,4968)} = 1.01$, $p = 0.4$, ANOVA). Subsequent inspection of data revealed that faces were slightly more relevant for the performance of our human subjects, but only in the saccade response task and not in the manual response task. These results indicate that the influence of object category on visual form effects for detection may be limited to overt gaze shifts.

Overt selection behavior is also facilitated by peripheral visual forms in monkeys

Since monkeys are an important animal model for investigating the neural mechanisms of spatial selection behaviors (Schall and Thompson, 1999; Reynolds and Chelazzi, 2004; Krauzlis et al., 2014; Basso and May, 2017), we next asked whether peripheral visual forms can have similar effects in these animals as in our human subjects. We used the same overt task design as in humans (see Materials and Methods), and we analyzed the monkeys' saccades. We confirmed that all 3 monkeys (Monkeys A, F, and M) performed the task correctly (% correct performance: 90.2 ± 2.1% SD, 92.8 ± 4% SD, and 84 ± 2.9% SD for Monkeys A, F, and M, respectively), and we also confirmed that individual monkeys' performance was significantly greater than chance in each of the 17 sessions collected across 3 monkeys (bootstrap test, $p < 0.001$). Following the same reasoning as in the human experiments, we compared saccadic response times on target congruent and target incongruent trials in each delay condition and monkey (Fig. 6). The results revealed two features that were

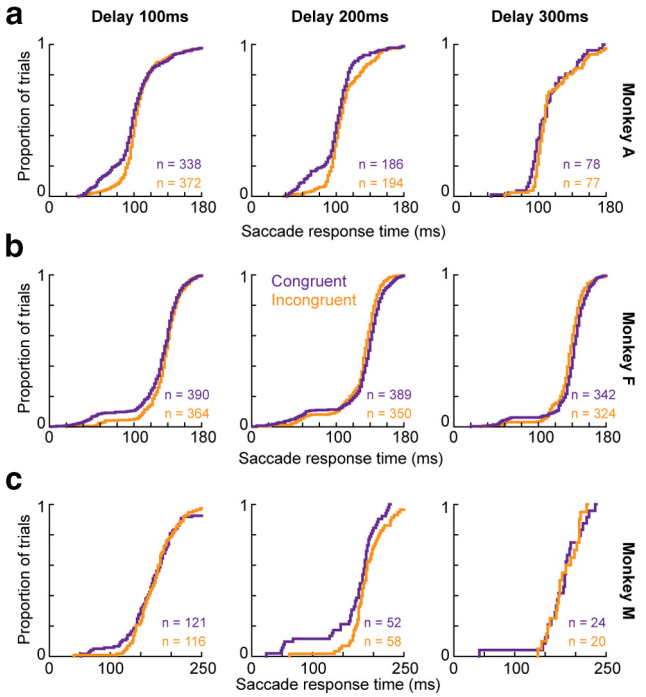

Figure 6. Distributions of saccade response times in the overt selection task in monkeys. ***a-c***, Cumulative distributions of saccade response times for target congruent and target incongruent trials in the overt selection task, similar to Figure 2*a*, *d*, in the example human subject. Each column represents a given delay condition (left, 100 ms; middle, 200 ms; right, 300 ms), and each row represents a monkey (A, F, and M, respectively). The number of samples in each cumulative distribution is shown in the corresponding panel. Color legend in the middle of the figure applies to all panels. All monkeys showed earlier response times for target congruent trials than for target incongruent trials when the response times were fast, but not when either the response times or target delay (300 ms) were long. This is a magnified effect of the human observations in Figure 2*f*.

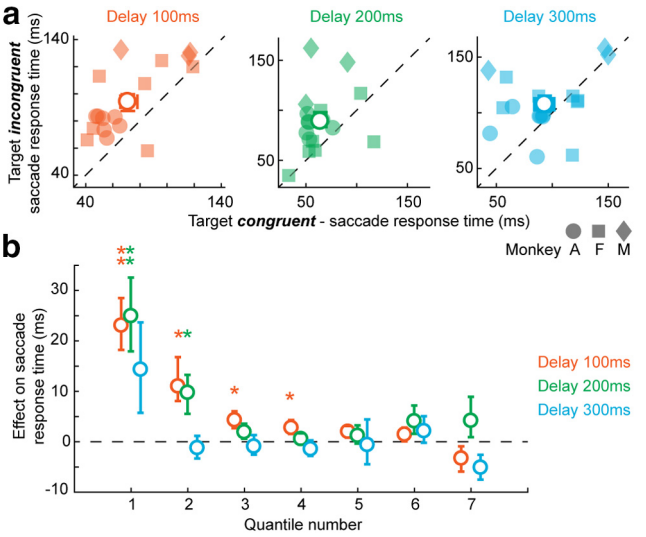

Figure 7. Visual form facilitation of early saccade responses in the overt selection task in monkeys. ***a***, Response times in each monkey were split into seven quantiles (see Materials and Methods). Each panel represents a paired comparison of median response times in the first quantile between target congruent and target incongruent trials for a fixed delay of 100 ms (red), 200 ms (green), or 300 ms (blue) across all sessions ($n = 17$; filled symbols) from 3 monkeys. Colored circle with errors bars in each panel represents mean ± SD across sessions. All monkeys showed faster response times for target congruent trials than target incongruent trials, consistent with our human results (Figure 2). ***b***, The effect of visual forms on saccade response times was quantified in each response time quantile as a difference in median response time between incongruent and congruent trials in the corresponding quantile. The mean effect on response times across sessions (colored circles) is shown separately for all seven quantiles and the three fixed delays. Error bars indicate SD across sessions. Statistical significance was tested using the Wilcoxon signed-rank test in the corresponding delay and quantile: $**p < 0.005$; $*p < 0.05$. Trials with fast response times were consistently associated with a facilitatory effect of visual forms on overt spatial selection behavior in monkeys.

consistent across all monkeys, and that were also consistent with our observations in humans when considering that monkey response times are generally faster than human response times. First, faster response times to congruent targets were limited to the early saccade responses as evident in the comparison between target congruent and target incongruent trials in the 100 ms delay condition (Fig. 6, first column of panels). Second, this facilitation of early saccade responses by visual forms was very weak in the 300 ms delay condition (Fig. 6, third column of panels) compared with the 100 ms delay condition.

We quantified this differential effect of visual forms on early and late saccade response times by splitting the response time distributions into 7 quantiles, such that the first quantile occupied the express-saccade part of the cumulative distributions in all conditions and monkeys. Express saccades represent a population of early saccades with very short latency, which appear to be distinct from the overall response time distribution (Fischer and Boch, 1983). Thus, in the cumulative distributions of response times (e.g., Fig. 6), the express-saccade part of response time distributions appears as an early distribution of trials before a plateau is reached in cumulative response time (i.e., an early tail in the global cumulative distribution). Paired comparisons of median response times for target congruent and target incongruent trials in the first quantile showed that saccadic responses were significantly faster for congruent targets in the 100 and 200 ms delay conditions (Fig. 7*a*; Wilcoxon signed-rank test, $p = 0.001$ in 100 ms, $p = 0.003$ in 200 ms) but not in the 300 ms delay condition (Fig. 7*a*; Wilcoxon signed-rank test, $p = 0.129$), consistent with the observations from Figure 6. The effect on all ranges (quantiles) of saccade response times in the three delay conditions is also shown in Figure 7*b*. As can be seen, there was a facilitatory effect of peripheral visual forms on saccade response times, but this was limited to the early saccadic responses and fell off abruptly for the 300 ms delay condition after the first quantile. The fall off was milder for the 100 and 200 ms delay conditions (Fig. 7*b*). These findings demonstrate that task-irrelevant visual forms facilitate early saccade responses in monkeys, and that this facilitation is the strongest in the first 200 ms of visual form processing, consistent with our earlier results in humans (Fig. 2*f*).

We next asked whether the weaker effect in the 300 ms delay condition was because of progressively lower trial counts in the longer delay conditions (Fig. 5; see Materials and Methods). We repeated the same analyses on all trials (with no exclusion of trials based on whether microsaccades occurred or not; see Materials and Methods). This resulted in comparable trial counts across the three delays (see Materials and Methods), and our results showed that median response times in the first quantile were significantly faster for congruent targets in the 100 and 200 ms delay conditions (Wilcoxon signed-rank test, $p = 0.0003$ in 100 ms, $p = 0.0003$ in 200 ms) but not in the 300 ms delay condition (Wilcoxon signed-rank test, $p = 0.58$). This further demonstrates that task-irrelevant visual forms influence early saccades within the first 200 ms of visual form processing.

Additionally, we performed ANOVAs on the first quantile response times with visual form, delay, and object categories as

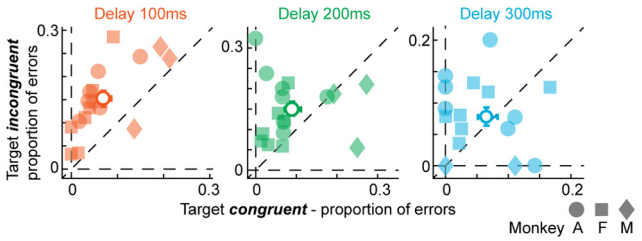

Figure 8. Visual form influence on errors in the overt spatial selection task in monkeys. Each panel represents a paired comparison of proportion of saccade response errors between target congruent and target incongruent trials for a fixed delay of 100 ms (red), 200 ms (green), or 300 ms (blue) across all sessions ($n = 17$; filled symbols) from 3 monkeys. Colored circle with errors bars in each panel represents mean ± SD across sessions. All monkeys had more errors in the target incongruent condition than in the target congruent condition, consistent with our human observations (Fig. 3).

factors, to test the influence of object categories on our findings (see Materials and Methods). The ANOVA results revealed a main effect of visual form factor ($F_{(1,489)} = 14.51$, $p = 0.0002$, ANOVA), consistent with our main findings (Fig. 7). We also found a significant interaction effect of visual form and category factors ($F_{(3,489)} = 3.47$, $p = 0.016$, ANOVA). Like in our human subjects above, this suggests that different categories of visual forms had different facilitatory effects on monkey spatial selection performance with saccades. Interestingly, the biggest facilitatory effects were with faces and fruits, and the least facilitatory effects were with inanimate objects. Nonetheless, we confirmed that inanimate objects, decidedly the least biologically relevant visual form category to monkeys in our experiments, still significantly influenced saccade response times within the first 200 ms of visual form processing (Wilcoxon signed-rank test, $p = 0.02$ in 100 ms, $p = 0.001$ in 200 ms, $p = 0.22$ in 300 ms). These findings suggest that object category plays a role in modulating the effect of visual forms on response times in monkeys, but that this influence is not necessarily limited by the ecological relevance of visual forms.

Visual forms capture more saccade errors in monkeys when incongruent with task requirements

Finally, in humans, the visual form facilitation of response times for target congruent trials was accompanied by the complimentary effect of more errors for targets that were incongruent with visual forms images (Fig. 3). We tested whether peripheral visual forms had this complementary effect on errors in monkeys as well. We found that monkeys indeed made significantly more errors when the saccade targets were incongruent with the visual forms compared with when the targets were congruent with the visual forms (Fig. 8). Interestingly, this effect on errors was strong and significant in the 100 ms (Wilcoxon signed-rank test, $p = 0.0006$) and 200 ms (Wilcoxon signed-rank test, $p = 0.02$) delay conditions but relatively weak and insignificant in the 300 ms delay condition (Wilcoxon signed-rank test, $p = 0.45$). This stronger effect on errors in the early delay conditions is consistent with the similar results from the saccade response times (Fig. 7). These findings demonstrate that peripheral visual forms capture more error saccades in well-trained monkeys, and that this effect on errors is the strongest in the first 200 ms of visual form processing.

Importantly, we also confirmed in monkeys that nonface visual forms still strongly influenced response times in the first quantile (Wilcoxon signed-rank test, $p = 0.008$ in 100 ms, $p = 0.006$ in 200 ms, $p = 0.18$ in 300 ms) and errors (Wilcoxon signed-rank test, $p = 0.0004$ in 100 ms, $p = 0.03$ in 200 ms, $p = 0.39$ in 300 ms) in the first 200 ms of visual form processing, as demonstrated in our main findings (Figs. 7*a*, 8).

Discussion

We investigated whether peripheral visual forms contribute to covert and overt spatial selection behaviors using a visual detection task in which visual forms were completely irrelevant. In humans, we found that visual forms facilitate the detection of spatially congruent targets with faster response times in both covert (Fig. 2*b*) and overt (Fig. 2*e*) tasks, and that this facilitation is evident in the first 100 ms of visual form processing. Importantly, visual forms incongruent with targets resulted in more errors in both covert (Fig. 3*a*) and overt (Fig. 3*b*) tasks, and this effect on errors was most pronounced after 200 ms of visual form processing. In addition, microsaccades before target presentation (but after visual form image onset) were biased toward visual forms in the covert task (Fig. 4*b*). Our results in monkeys revealed a similar pattern of visual form effects seen in humans with two notable differences. First, visual form facilitation of response times was specific to early saccadic responses (Fig. 7*b*), likely because monkey saccadic response times are faster than those of humans. Second, the visual form effects on response times and errors were limited to the first 200 ms of visual form processing (Figs. 7, 8). Overall, these findings demonstrate that peripheral visual forms contribute to covert and overt spatial selection in ways that resemble the effects of spatial cues on orienting behaviors (Posner, 1980).

Low-level visual factors cannot explain visual form influences on response times

Low-level visual factors related to luminance, spatial frequency content, and target contrast modulate neuronal activity in visual and sensorimotor regions of the brain (Ohayon et al., 2012; Chen and Hafed, 2018; Chen et al., 2018; Vinke and Ling, 2020), and therefore may influence behavioral responses. For this reason, we took several measures in the design of the image and target stimuli to minimize the contribution of low-level visual factors to our findings, particularly on response times. First, we equalized all visual form images and their corresponding 50% phase-scrambled images iteratively for luminance distributions and the Fourier spectra (see Materials and Methods). Second, we chose the target to be of the highest contrast and adjusted the size so that target detection, and hence the perceived contrast of the target, was suprathreshold for both visual form and phase-scrambled image backgrounds (see Materials and Methods). Thus, we suggest that low-level visual factors were unlikely to have influenced our results showing visual form facilitation of response times.

High-level cognitive factors cannot explain visual form influence on response times

Cognitive factors related to behavioral relevance, novelty, and reward also modulate neuronal activity in visuomotor brain regions, such as the SC, and hence can shape orienting behaviors (Basso and Wurtz, 1997; Ikeda and Hikosaka, 2003; Boehnke et al., 2011; Herman and Krauzlis, 2017). These factors are again unlikely to have influenced our findings for the following reasons. First, we made the visual form and the phase-scrambled images completely irrelevant to behavior in both covert and overt tasks, and in both humans and monkeys. Second, the same subjects participated in both covert and overt tasks that used the same images (see Materials and Methods). In addition, all

monkeys were trained with the same images in at least 7 training sessions before the experimental sessions. Third, none of the images was associated with reward, in humans and monkeys, as they were irrelevant to the performance in the task. Thus, cognitive factors related to behavioral relevance, novelty, and reward were unlikely to have influenced our findings showing visual form facilitation of response times.

Face stimuli alone cannot account for visual form influence on response times

Faces are of ecological value, and the influence of faces on goal-directed and free-viewing saccade behaviors is well documented (Bindemann et al., 2007; Xu-Wilson et al., 2009; Morand et al., 2010; Devue et al., 2012; Boucart et al., 2016; Kauffmann et al., 2019; Buonocore et al., 2020). Importantly, there is growing evidence that faces are rapidly processed through a network of subcortical structures, including the SC (Johnson, 2005; Nguyen et al., 2016; Le et al., 2020), which also plays a crucial role in spatial selection (McPeek and Keller, 2004; Lovejoy and Krauzlis, 2010). To confirm that face images alone did not disproportionately contribute to our results, we repeated all of our analyses of covert and overt tasks in humans by excluding trials with face images. Results showed that response times were equally strongly affected by nonface visual forms alone in both covert (Fig. 5*a*) and overt tasks (Fig. 5*b*). Importantly, we also observed significant biases in microsaccades to nonface visual forms (Fig. 5*c*). Additionally, we also confirmed in monkeys that nonface visual forms strongly influenced the response times. These control analyses show that face stimuli alone cannot account for visual form influence on response times in both covert and overt tasks, and most importantly, demonstrate that all visual forms can influence spatial selection. Nonetheless, it would be interesting in the future to identify potential graded influences of different visual form categories on spatial selection performance. For example, our ANOVAs did show stronger effects of fruits and faces on monkey saccade performance than inanimate objects. This suggests that ecological relevance, even in the oculomotor system, needs to be considered when interpreting neural and behavioral effects. Indeed, increasing evidence supports the idea of an oculomotor system organization that is in line with the image statistics of the environment in which we make eye movements (Hafed and Chen, 2016; Chen et al., 2018).

Comparison of visual form effects in humans and monkeys

A comparison of visual form effects in humans and monkeys during the overt task revealed interesting species differences. For example, visual form effects on response times in monkeys were confined to the earliest saccades (Fig. 7*b*), unlike in humans where a similar analysis revealed visual form effects across all quantiles in all delay conditions (Fig. 2*b*). Similarly, visual form effects on errors in monkeys were more pronounced in the early delay conditions (100 and 200 ms; see Fig. 8) with a weaker effect in the late 300 ms delay condition, unlike in humans where this pattern was almost reversed; effects on errors were the weakest in the 100 ms condition (Fig. 3*b*). We suggest that the predominance of visual form effects on earliest responses and delay conditions in monkeys may be related to their behavioral training. Specifically, these were highly trained animals with short saccadic response times in general. With the longer delay periods (e.g., 200 and 300 ms), these delay periods were often much longer than the actual saccadic reaction times that would have been elicited to the visual form images themselves (e.g., see Fig. 7*a*). The long delay periods therefore required actively suppressing saccades to properly receive rewards in the task, which eliminated the visual form effects that still occurred automatically with shorter latencies. Indeed, the monkeys' final reaction times on successful trials were much shorter than those of the human subjects in the same task (compare Fig. 2 with Fig. 7).

Visual-form based selection differs from object-based attention

Space-based or spatial attention refers to behavioral benefits conferred by spatial cues exclusively at the cued location (Carrasco, 2011). In object-based attention, the cueing benefits extend to all spatial locations occupied by the object at the cued location (Duncan, 1984; Egly et al., 1994; Abrams and Law, 2000). Our demonstration of spatial selection based on visual forms is different from object-based attention because there were no explicit spatial cues in our task, and, most importantly, the visual forms were irrelevant in our task. However, it is very likely that both object-based and visual form-based selection mechanisms involve common visual processes related to segmentation and perceptual grouping (Driver et al., 2001; Baldauf and Desimone, 2014), and may operate outside of the modulation of sensory processing mechanisms associated with spatial attention (Shomstein and Yantis, 2002; Reynolds and Chelazzi, 2004; Chou et al., 2014; but see Roelfsema et al., 1998).

Neural circuits representing visual-form based spatial maps for orienting

The influence of peripheral visual forms on target detection as early as 100 ms suggests a neural circuit that rapidly links visual form processing with spatial maps in sensorimotor structures, such as the SC (Robinson, 1972; Chen et al., 2019). Recent evidence in a new region of the primate temporal cortex shows rapid object selectivity and detection-related signals that were causally dependent on midbrain SC activity (Bogadhi et al., 2019b). Based on this evidence, we suspect that SC neurons might signal peripheral visual forms and bias spatial selection. Recent findings in monkeys and mice further demonstrate the visual capabilities of SC neurons in representing visual statistics and properties of the natural environment that are innately relevant to our behaviors (Hafed and Chen, 2016; Chen et al., 2018; Lee et al., 2020).

Of course, visual form recognition is also accomplished in the primate inferotemporal cortex through feedforward visual cortical circuits. This can possibly influence sensorimotor structures, such as the SC, for spatial selection through direct projections (Cerkevich et al., 2014). However, the time course of visual form recognition in the traditional inferotemporal regions is not entirely consistent with our results showing rapid visual form facilitation (Kreiman et al., 2006; Tsao et al., 2006). Therefore, we suggest that a circuit linking SC with the temporal cortex, possibly through pulvinar or amygdala, may be at play in linking rapid visual form recognition with spatial selection (Harting et al., 1991; Boussaoud et al., 1992; Hadj-Bouziane et al., 2012; Rafal et al., 2015; Soares et al., 2017). Future studies investigating the subcortical and cortical contributions to visual form recognition, particularly in the periphery, will identify the candidate circuit mediating the visual form influence on spatial selection.

References

Abrams RA, Law MB (2000) Object-based visual attention with endogenous orienting. Percept Psychophys 62:818–833.

Awh E, Belopolsky AV, Theeuwes J (2012) Top-down versus bottom-up attentional control: a failed theoretical dichotomy. Trends Cogn Sci 16:437–443.

Baldauf D, Desimone R (2014) Neural mechanisms of object-based attention. Science 344:424–427.

Basso MA, Wurtz RH (1997) Modulation of neuronal activity by target uncertainty. Nature 389:66–69.

Basso MA, May PJ (2017) Circuits for action and cognition: a view from the superior colliculus. Annu Rev Vis Sci 3:197–226.

Baumeler D, Schönhammer JG, Born S (2020) Microsaccade dynamics in the attentional repulsion effect. Vision Res 170:46–52.

Bindemann M, Burton AM, Langton SR, Schweinberger SR, Doherty MJ (2007) The control of attention to faces. J Vis 7:15.1–8.

Bisley JW, Goldberg ME (2003) Neuronal activity in the lateral intraparietal area and spatial attention. Science 299:81–86.

Bisley JW, Mirpour K (2019) The neural instantiation of a priority map. Curr Opin Psychol 29:108–112.

Boehnke SE, Berg DJ, Marino RA, Baldi PF, Itti L, Munoz DP (2011) Visual adaptation and novelty responses in the superior colliculus. Eur J Neurosci 34:766–779.

Bogadhi A, Bollimunta A, Leopold D, Krauzlis R (2019a) Spatial attention deficits are causally linked to an area in macaque temporal cortex. Curr Biol 29:726–736.e4.

Bogadhi AR, Katz LN, Bollimunta A, Leopold DA, Krauzlis RJ (2019b) Midbrain activity supports high-level visual properties in primate temporal cortex. Biorxiv. doi: https://doi.org/10.1101/841155.

Boucart M, Lenoble Q, Quettelart J, Szaffarczyk S, Despretz P, Thorpe SJ (2016) Finding faces, animals, and vehicles in far peripheral vision. J Vis 16:10.

Boussaoud D, Desimone R, Ungerleider LG (1992) Subcortical connections of visual areas MST and FST in macaques. Vis Neurosci 9:291–302.

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

Buonocore A, Chen CY, Tian X, Idrees S, Münch TA, Hafed ZM (2017) Alteration of the microsaccadic velocity-amplitude main sequence relationship after visual transients: implications for models of saccade control. J Neurophysiol 117:1894–1910.

Buonocore A, Dimigen O, Melcher D (2020) Post-saccadic face processing is modulated by pre-saccadic preview: evidence from fixation-related potentials. J Neurosci 40:2305–2313.

Carrasco M (2011) Visual attention: the past 25 years. Vision Res 51:1484–1525.

Cerkevich C, Lyon D, Balaram P, Kaas J (2014) Distribution of cortical neurons projecting to the superior colliculus in macaque monkeys. Eye Brain 2014:121–137. Volume

Chelazzi L, Eštočinová J, Calletti R, Gerfo EL, Sani I, Libera CD, Santandrea E (2014) Altering spatial priority maps via reward-based learning. J Neurosci 34:8594–8604.

Chen CY, Hafed ZM (2013) Postmicrosaccadic enhancement of slow eye movements. J Neurosci 33:5375–5386.

Chen CY, Hafed ZM (2018) Orientation and contrast tuning properties and temporal flicker fusion characteristics of primate superior colliculus neurons. Front Neural Circuits 12:58.

Chen CY, Ignashchenkova A, Thier P, Hafed ZM (2015) Neuronal response gain enhancement prior to microsaccades. Curr Biol 25:2065–2074.

Chen CY, Sonnenberg L, Weller S, Witschel T, Hafed ZM (2018) Spatial frequency sensitivity in macaque midbrain. Nat Commun 9:2852.

Chen CY, Hoffmann KP, Distler C, Hafed ZM (2019) The foveal visual representation of the primate superior colliculus. Curr Biol 29:2109–2119.e7.

Chen X, Zirnsak M, Vega GM, Govil E, Lomber SG, Moore T (2020) Parietal cortex regulates visual salience and salience-driven behavior. Neuron 106:177–187.e4.

Chou WL, Yeh SL, Chen CC (2014) Distinct mechanisms subserve location- and object-based visual attention. Front Psychol 5:456.

Crouzet SM, Thorpe SJ (2011) Low-level cues and ultra-fast face detection. Front Psychol 2:342.

Crouzet SM, Kirchner H, Thorpe SJ (2010) Fast saccades toward faces: face detection in just 100 ms. J Vis 10:16.1-17.

Devue C, Belopolsky AV, Theeuwes J (2012) Oculomotor guidance and capture by irrelevant faces. PLoS One 7:e34598.

Drewes J, Trommershäuser J, Gegenfurtner KR (2011) Parallel visual search and rapid animal detection in natural scenes. J Vis 11:20.

Driver J, Davis G, Russell C, Turatto M, Freeman E (2001) Segmentation, attention and phenomenal visual objects. Cognition 80:61–95.

Duncan J (1984) Selective attention and the organization of visual information. J Exp Psychol Gen 113:501–517.

Eastman KM, Huk AC (2012) PLDAPS: a hardware architecture and software toolbox for neurophysiology requiring complex visual stimuli and online behavioral control. Front Neuroinform 6:1.

Egly R, Driver J, Rafal RD (1994) Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. J Exp Psychol Gen 123:161–177.

Einhäuser W, Spain M, Perona P (2008) Objects predict fixations better than early saliency. J Vis 8:18.

Engbert R, Kliegl R (2003) Microsaccades uncover the orientation of covert attention. Vision Res 43:1035–1045.

Fecteau JH, Munoz DP (2006) Salience, relevance, and firing: a priority map for target selection. Trends Cogn Sci 10:382–390.

Fischer B, Boch R (1983) Saccadic eye movements after extremely short reaction times in the monkey. Brain Res 260:21–26.

Gottlieb JP, Kusunoki M, Goldberg ME (1998) The representation of visual salience in monkey parietal cortex. Nature 391:481–484.

Hadj-Bouziane F, Liu N, Bell AH, Gothard KM, Luh WM, Tootell RB, Murray EA, Ungerleider LG (2012) Amygdala lesions disrupt modulation of functional MRI activity evoked by facial expression in the monkey inferior temporal cortex. Proc Natl Acad Sci USA 109:E3640–E3648.

Hafed ZM (2013) Alteration of visual perception prior to microsaccades. Neuron 77:775–786.

Hafed ZM, Clark JJ (2002) Microsaccades as an overt measure of covert attention shifts. Vision Res 42:2533–2545.

Hafed ZM, Chen CY (2016) Sharper, stronger, faster upper visual field representation in primate superior colliculus. Curr Biol 26:1647–1658.

Harting J, Huerta M, Hashikawa T, van Lieshout D (1991) Projection of the mammalian superior colliculus upon the dorsal lateral geniculate nucleus: organization of tectogeniculate pathways in nineteen species. J Comp Neurol 304:275–306.

Herman JP, Krauzlis RJ (2017) Color-change detection activity in the primate superior colliculus. eNeuro 4:0046-17.2017.

Honey C, Kirchner H, VanRullen R (2008) Faces in the cloud: Fourier power spectrum biases ultrarapid face detection. J Vis 8:9.

Ignashchenkova A, Dicke PW, Haarmeier T, Thier P (2004) Neuron-specific contribution of the superior colliculus to overt and covert shifts of attention. Nat Neurosci 7:56–64.

Ikeda T, Hikosaka O (2003) Reward-dependent gain and bias of visual responses in primate superior colliculus. Neuron 39:693–700.

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

Johnson MH (2005) Subcortical face processing. Nat Rev Neurosci 6:766–774.

Kauffmann L, Peyrin C, Chauvin A, Entzmann L, Breuil C, Guyader N (2019) Face perception influences the programming of eye movements. Sci Rep 9:560.

Kirchner H, Thorpe SJ (2006) Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. Vision Res 46:1762–1776.

Krauzlis R, Miles F (1996) Release of fixation for pursuit and saccades in humans: evidence for shared inputs acting on different neural substrates. J Neurophysiol 76:2822–2833.

Krauzlis RJ, Bollimunta A, Arcizet F, Wang L (2014) Attention as an effect not a cause. Trends Cogn Sci 18:457–464.

Kreiman G, Hung CP, Kraskov A, Quiroga RQ, Poggio T, DiCarlo JJ (2006) Object selectivity of local field potentials and spikes in the macaque inferior temporal cortex. Neuron 49:433–445.

Kümmerer M, Theis L, Bethge M (2014) Deep gaze I: boosting saliency prediction with feature maps trained on ImageNet. arXiv

Le QV, Le QV, Nishimaru H, Matsumoto J, Takamura Y, Hori E, Maior RS, Tomaz C, Ono T, Nishijo H (2020) A prototypical template for rapid face detection is embedded in the monkey superior colliculus. Frontiers Syst Neurosci 14:5.

Lee KH, Tran A, Turan Z, Meister M (2020) The sifting of visual information in the superior colliculus. Elife 9:e50678.

Lovejoy LP, Krauzlis RJ (2010) Inactivation of primate superior colliculus impairs covert selection of signals for perceptual judgments. Nat Neurosci 13:261–266.

McPeek RM, Keller EL (2004) Deficits in saccade target selection after inactivation of superior colliculus. Nat Neurosci 7:757–763.
Morand SM, Grosbras MH, Caldara R, Harvey M (2010) Looking away from faces: influence of high-level visual processes on saccade programming. J Vis 10:16.1-10.
Nguyen MN, Nishimaru H, Matsumoto J, Le QV, Hori E, Maior RS, Tomaz C, Ono T, Nishijo H (2016) Population coding of facial information in the monkey superior colliculus and pulvinar. Front Neurosci 10:583.
Ohayon S, Freiwald WA, Tsao DY (2012) What makes a cell face selective? The importance of contrast. Neuron 74:567–581.
Pastukhov A, Braun J (2010) Rare but precious: microsaccades are highly informative about attentional allocation. Vision Res 50:1173–1184.
Pastukhov A, Vonau V, Stonkute S, Braun J (2013) Spatial and temporal attention revealed by microsaccades. Vision Res 85:45–57.
Peel TR, Hafed ZM, Dash S, Lomber SG, Corneil BD (2016) A causal role for the cortical frontal eye fields in microsaccade deployment. Plos Biol 14: e1002531.
Posner MI (1980) Orienting of attention. Q J Exp Psychol 32:3–25.
Rafal RD, Koller K, Bultitude JH, Mullins P, Ward R, Mitchell AS, Bell AH (2015) Connectivity between the superior colliculus and the amygdala in humans and macaque monkeys: virtual dissection with probabilistic DTI tractography. J Neurophysiol 114:1947–1962.
Reynolds JH, Chelazzi L (2004) Attentional modulation of visual processing. Annu Rev Neurosci 27:611–647.
Robinson D (1972) Eye movements evoked by collicular stimulation in the alert monkey. Vision Res 12:1795–1808.
Roelfsema PR, Lamme VA, Spekreijse H (1998) Object-based attention in the primary visual cortex of the macaque monkey. Nature 395:376–381.
Rolfs M, Kliegl R, Engbert R (2008) Toward a model of microsaccade generation: the case of microsaccadic inhibition. J Vision 8:5.
Sapountzis P, Paneri S, Gregoriou GG (2018) Distinct roles of prefrontal and parietal areas in the encoding of attentional priority. Proc Natl Acad Sci USA 115:201804643.
Schall JD, Thompson KG (1999) Neural selection and control of visually guided eye movements. Annu Rev Neurosci 22:241–259.
Shomstein S, Yantis S (2002) Object-based attention: sensory modulation or priority setting? Percept Psychophys 64:41–51.
Skinner J, Buonocore A, Hafed ZM (2019) Transfer function of the rhesus macaque oculomotor system for small-amplitude slow motion trajectories. J Neurophysiol 121:513–529.
Soares SC, Maior RS, Isbell LA, Tomaz C, Nishijo H (2017) Fast detector/first responder: interactions between the superior colliculus-pulvinar pathway and stimuli relevant to primates. Front Neurosci 11:67.
Sprague TC, Itthipuripat S, Vo VA, Serences JT (2018) Dissociable signatures of visual salience and behavioral relevance across attentional priority maps in human cortex. J Neurophysiol 119:2153–2165.
Stemmann H, Freiwald WA (2019) Evidence for an attentional priority map in inferotemporal cortex. Proc Natl Acad Sci USA 116:23797–23805.
Tian X, Yoshida M, Hafed ZM (2016) A microsaccadic account of attentional capture and inhibition of return in Posner cueing. Front Syst Neurosci 10:23.
Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. Science 311:670–674.
Veale R, Hafed ZM, Yoshida M (2017) How is visual salience computed in the brain? Insights from behaviour, neurobiology and modelling. Philos Trans R Soc Lond B Biol Sci 372:20160113.
Vinke LN, Ling S (2020) Luminance potentiates human visuocortical responses. J Neurophysiol 123:473–483.
White BJ, Kan JY, Levy R, Itti L, Munoz DP (2017) Superior colliculus encodes visual saliency before the primary visual cortex. Proc Natl Acad Sci USA 114:9451–9456.
Wichmann FA, Drewes J, Rosas P, Gegenfurtner KR (2010) Animal detection in natural scenes: critical features revisited. J Vis 10:6.1-27.
Willenbockel V, Sadr J, Fiset D, Horne GO, Gosselin F, Tanaka JW (2010) Controlling low-level image properties: the SHINE toolbox. Behav Res Methods 42:671–684.
Xu-Wilson M, Zee DS, Shadmehr R (2009) The intrinsic value of visual information affects saccade velocities. Exp Brain Res 196:475–481.
Yanulevskaya V, Uijlings J, Geusebroek JM, Sebe N, Smeulders A (2013) A proto-object-based computational model for visual saliency. J Vis 13:27.
Yan Y, Zhaoping L, Li W (2018) Bottom-up saliency and top-down learning in the primary visual cortex of monkeys. Proc Natl Acad Sci USA 115:10499–10504.
Zhu W, Drewes J, Gegenfurtner KR (2013) Animal detection in natural images: effects of color and image database. PLoS One 8:e75816.