Object-based attention: the effect of selective and divided attention

Zachary Raymond Ernst

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of philosophy

University of Washington

2012

Reading Committee:

Geoffrey M. Boynton, Chair

John Palmer

Scott Murray

Anitha Pasupathy

Program Authorized to Offer Degree:

Psychology

CONTENTS

ACKNOWLEDGEMENTS

I am indebted to my parents who put up with my incessant demands to know why, and my intransigent refusal to accept, “I don’t know.” As academics, they instilled in me great curiosity and wonder for the natural world. It is rare to refer to ones parents as friends, but there is no better way to characterize our relationship than to that of best friends. Some people fight hard to be nothing like their parents, but I celebrate the heredity of my dispositions with pride. Along with my parents is my sister Jessica, who I’ve always looked up to, and who blazed the path to a Ph. D. before me. I walk in her footsteps motivated not by sibling rivalry but by admiration and a desire to be as learned as she.

In addition to a supportive and loving family, the friends that I have made at each step along the way have contributed to the person I have become and still aspire to be. Of particular note are my high school friends Ross Gladwin and Trevor Zhou; college friends Sarahmaria Gomez, Nate DeYoung and Aaron Holzman; and Seattle friends Alec Scharff, Alexander Pagliere and Chie Kawahara. Finally, I would be lost in many ways without my partner and companion Hilary Hammell, who believes in me even when my own self confidence is on edge, and who engenders in me the indescribable feeling of being both utterly secure and entirely free.

None of this work, or the education behind it, would have been possible without the mentorship I received at the University of Washington. First and foremost, I want to thank my advisor Geoff Boynton, who took a risk by offering me a position in his lab despite my lack of experience in behavioral psychology. Geoff’s lab is built around mutual respect and he treats his students as colleagues and friends. He is an excellent teacher and I owe much of my programming and analytical expertise to him. Second, I would like to thank John Palmer who took a

special interest in my work late in my graduate career. He quickly assumed the role of a mentor and took the initiative in setting up weekly meetings to discuss my work and foment a collaboration that resulted in a large portion of this dissertation. John is a paragon of science by way of his principles and mentorship. I regret not having the opportunity to collaborate with Scott as much as I would have liked. Scott always provided good advice, scientific and otherwise, and his thoughtfulness always speaks volumes. Outside of my committee, I would like to thank Mehrdad Jazayeri who came to Geoff with the kernel of an experiment that I was lucky enough to become involved with and take on as my own project. No step in the process from the design, analysis, or writing was without Mehrdad's copious input.

This dissertation focuses on two projects that I have prepared and submitted to peer reviewed journals, but as a whole represents only the tip of the iceberg of the work that I have done here at the UW. Of note was the cross-modal project that I began upon my arrival to Geoff's lab. Using psychophysical methods, I measured the spread of attention between the visual and auditory modalities as a function of their perceptual congruency. Although this work did not culminate in a publication, I presented the results in my first year talk to the department and as a poster at the annual Vision Science Society meeting. As an introduction to behavioral psychology, Geoff would have been hard pressed to have chosen a better introduction to the field. Second, when the department invested in an EEG system, I collaborated with Mark Pettet on designing, collecting, and analyzing a steady-state visual evoked potential experiment on the dynamics of object-based attention. I learned a lot from Mark about signal processing and am grateful for Geoff's for allowing me to explore my divergent interests and to always provide encouragement when I needed it most.

STUDYING OBJECT-BASED ATTENTION

Introduction

The sensory organs slavishly transmit all of the information that they collect to the cortex. But only a subset of the information that is sensed is inevitably perceived. This phenomenon reflects two key principles of perception: 1) perception is capacity limited (Broadbent, 1958; Neisser, 1967; Treisman, 1960); and 2) attention provides a flexible means of prioritizing some sensory information at the expense of other sensory information (Duncan, 1980; Posner, Snyder, & Davidson, 1980; Treisman, 1964a, 1964b). This dissertation is concerned first with whether selection is feature- or object-based, and second, with the capacity limits of divided attention for simple features like an object's color and direction of motion.

The study of visual attention has historically focused on either characterizing the nature of selective attention, or identifying the capacity limits of divided attention. The former is concerned with the means by which visual information is selected, and the effect that selection has on perception and the underlying neural response. The visual system flexibly allocates attention to meet the demands of specific behavioral goals. Visual information can be selected on the basis of spatial location (Eriksen & St. James, 1986; Posner et al., 1980), as when a free-throw shooter must focus on the hoop and ignore the crowd. Or, a particular feature can be selected independent of spatial location (Saenz, Buracas, & Boynton, 2002; Treue & Martinez-Trujillo, 1999), as when looking for an red frisbee amidst the overgrown green grass. Or, a collection of features can be selected in order to segment an object from its surroundings (Duncan, 1984; Neisser, 1967), as does the tennis umpire when tracking the moving yellow ball in order to call the shot in or out.

Divided attention, on the other hand, is concerned with the perceptual capacity to attend to multiple things at the same time (Broadbent, 1958;

Kahneman, 1973; Scharff & Palmer, 2008; Townsend, 1990). In some situations perception has unlimited capacity (Gardner, 1973). For example, spotting a single office light shining through the window of a skyscraper at night is easy; regardless of how many floors of dark windows the skyscraper has, the lit window appears to *pop out*. In other situations capacity is limited (Palmer, 1990; Shaw, 1980; Sternberg, 1969). For example, in scanning a dissertation for misspelled words, the longer the dissertation, the longer it will take to scan every word.

My research focuses on two aspects of object-based attention. First, I explore a situation in which the predictions of feature- and object-based attention conflict. Namely, can a single feature of an object be selected or are all the features of an object selected in concert even if only one feature is behaviorally relevant? I show evidence for the latter, which suggests that object-based selection supersedes features-based selection. These results motivated the following question—tested in the second half of this dissertation—if the features of an object are selected holistically, then will there be no perceptual cost in dividing attention across features within the same object? Furthermore, if object perception is limited in capacity, then dividing attention across features that belong to separate objects may be more difficult than attending to one at a time.

Selective attention

Selective attention research is primarily concerned with the physiological response to, and perception of, the selected versus non-selected information. The *partially valid cueing paradigm*, introduced by Posner (Posner, 1980), has been extensively used to quantify the effects of selective attention. In this paradigm, an observer is presented multiple stimuli on a given trial and then asked to report some perceptual aspect of one of the stimuli. Critically, the observer is cued to attend to a particular stimulus at the start of each trial. In order to compare the perceptual quality of the attended and unattended stimulus, the uncued stimulus is queried on some lesser proportion of trials. Because the cue contains valid information as to which stimulus is more likely to be queried, it is assumed that the observer prioritizes the cued stimulus over the uncued stimulus. The difference in behavioral performance, or neural response, between the cued and uncued stimulus serves to quantify the effect of selective attention.

The validity of the partially valid cueing paradigm depends on the assumption that the uncued stimulus is in fact unattended despite being behaviorally relevant (albeit in a lesser proportion of trials). Measuring the perceptual fate of an unattended stimulus is challenging because the experimenter faces the conflicting interest of ensuring that the subject does not attend to the very thing they wish to question. Although psychophysical tools have been developed to circumvent this issue, I will instead focus on how functional magnetic resonance imaging (fMRI) can covertly measure the response to a stimulus that the observer is told to ignore and is never asked about. Using a *fully valid cue*, fMRI provides a noninvasive means of measuring the response to an identical stimulus when it is relevant versus when it is irrelevant to an observer's task.

Functional MRI is naturally suited to measure the effects of spatial attention. Due to the retinotopic organization of visual cortex, spatially distinct stimuli will generate responses in different regions of cortex. By analyzing the associated regions of interest, the response to the cued and uncued stimulus can be simultaneously measured. This work provided surprising evidence that visual attention modulates responses as early as primary visual cortex (Gandhi, Heeger, & Boynton, 1999; Watanabe et al., 1998), and even increases baseline activity in the absence of visual stimulation (Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999).

Functional MRI also led the field in the study of feature-based attention. fMRI revealed that directing attention to a particular feature, such as a color or direction of motion, modulates the response to like features across visual field (Saenz et al., 2002; Serences & Boynton, 2007) and independent of spatial attention (Boynton, 2005a; Zhang & Luck, 2008).

Finally, fMRI has also be used to investigate whether object-based selection occurs even when only a single feature of an object is behaviorally relevant

(O'Craven, Downing, & Kanwisher, 1999). Exploiting the functional specialization of specific cortical areas for motion, faces, and places, O'Craven et al. measured an enhanced response to a task-irrelevant feature (motion or face or house) while the subject performed a task on the other feature (object identity or motion). These results suggest that all of an object's features are selected.

In the first half of this dissertation I describe the physiological effect of selective attention on the activity of feature-selective neurons in visual cortex. Critically, the task involved attending to a particular feature of object composed of multiple features. If selection can be restricted to a particular feature, as predicted by feature-based attention, than the hemodynamic response (measured by fMRI) should reflect which feature was cued. However, if all of a relevant object's features are selected, as predicted by object-based attention, then the hemodynamic response to each of the object's features should be modulated regardless of which feature was cued. I provide evidence for the latter, consistent with the prediction of object-based selection. Following from this result is the question addressed in the second half of this dissertation—what are the consequences of object-based selection when dividing attention across features within an object?

Divided attention

Research on divided attention is concerned with the ability to attend to multiple sources of information at the same time. Multiple psychophysical approaches have been developed to study this phenomenon. The two primary methods are the *set-size* and the *dual-task* paradigms. In a typical set-size experiment (Treisman & Gelade, 1980) an observer's behavioral performance (e.g. detecting a red target letter) is plotted as a function of the set-size, the amount of distracting information (e.g. the number of irrelevant green distracter letters in the display). If performance remains constant as the set-size increases than it is reasoned that the observer has an unlimited capacity to perform the given task. However, if performance decreases with set-size then it is argued that the particular task is capacity limited. Unfortunately, these conclusions require certain assumption regarding internal noise and decision error (Palmer, 1995), which complicate the interpretation. In the dual-task paradigm, the display size and stimulus are held constant and instead the number of perceptual tasks are varied (Sperling & Melchner, 1978). Again, constant performance across single and dual-task conditions is associated with an unlimited capacity process, where as a dual-task deficit is associated with a limited capacity process.

The results of the fMRI experiment led us to speculate on the capacity limits underlying the processing of multiple features within and between objects. Because the attentional modulation of the hemodynamic response was equivalent across features within an object, we hypothesized that attending to multiple features within an object should reflect an unlimited capacity to process (Blaser, Pylyshyn, & Holcombe, 2000; Bonnel & Prinzmetal, 1998; Duncan, 1984).

To test this prediction, we employed a dual-task paradigm in which attention was either divided across features within an object or across features between objects. We observed no cost in dual-task performance when attention was divided across features within an object relative to a large cost when attention was divided between objects. These results support our physiological observations that object-based attention selects all features within a surface, and suggests that monitoring multiple features within an object is an unlimited capacity process. In addition, our behavioral results reveal that that attending to multiple objects is capacity limited, and suggests that objects might impose a perceptual bottleneck (Scharff, Palmer, & Moore, 2011).

THE SPREAD OF ATTENTION ACROSS FEATURES OF A SURFACE

In collaboration with:
Mehrdad Jazayeri & Geoffrey M. Boynton
(In preparation)

Contrasting theories of visual attention have emphasized selection by spatial location, individual features and whole objects. We used functional magnetic resonance imaging (fMRI) to ask whether and how attention to one feature of an object spreads to other features of the same object. Subjects viewed two spatially superimposed surfaces of random dots that were segregated by distinct color-motion conjunctions. The color and direction of motion of each surface changed smoothly and in a cyclical fashion. Subjects were required to track one feature (e.g., color) of one of the two surfaces and detect brief moments when the attended feature diverged from its smooth trajectory. To tease apart the effect of attention to individual features on the hemodynamic response, we used a *frequency-tagging* scheme. In this scheme, the stimulus features (color and direction of motion) are modulated periodically at distinct frequencies so that the contribution of each feature to the hemodynamics can be inferred from harmonic responses at the corresponding frequency. We found that attention to one feature (e.g., color) of one surface not only increased the response modulation to the attended feature but also to the other feature (e.g., motion) of the same surface. This attentional modulation was evident in multiple visual areas and was present as early as V1. The spread of attention to the behaviorally irrelevant features of a surface suggests that attention might automatically select all features of a single object. Accordingly, we found that tracking both features of a single surface simultaneously did not incur any additional cost in behavioral performance. These results suggest that object-based attention in vision is supported by an enhancement of feature-specific sensory signals in the visual cortex.

INTRODUCTION

Selective attention improves information processing for a subset of relevant stimuli, usually at the expense of irrelevant stimuli. Attention can select a region of space (spatial attention), a stimulus feature (feature attention), or a whole object (object attention). What distinguishes object- and feature-based attention is that object-based attention improves processing of all features of a selected object with little or no additional cost. For example, when asked to monitor multiple features simultaneously, subjects are more accurate when the attended features belong to the same object compared to when they are from different objects (Duncan, 1984; Blaser et al., 2000; Rodriguez et al., 2002). Despite its importance in behavior, little is known about the mechanisms by which object-based attention influences the representation of individual features in the brain.

A common challenge in studying the mechanisms of object-based attention in humans is that existing tools such as functional magnetic resonance imaging (fMRI) do not have the requisite resolution to tease apart the representation of individual features when they overlap in space and time. Pattern classification techniques have provided a means to circumvent this problem, by extracting information from the pattern of hemodynamic (BOLD) responses across voxels (Boynton, 2005b). For example, in occipital cortex, pattern classification can extract information about orientation (Kamitani and Tong, 2005), directions of motion (Kamitani & Tong, 2006), and color (Brouwer & Heeger, 2009; Kamitani & Tong, 2005, 2006). Moreover, this technique has been used to demonstrate how attention to a specific feature can selectively and reliably modulate the pattern of fMRI responses to that feature (Kamitani & Tong, 2005, 2006; Serences & Boynton, 2007).

Pattern classification methods use sophisticated algorithms to *decode* information that is not immediately accessible at the level of the spatially averaged BOLD signal. An alternative to this strategy, and one that we have used here, is to design stimuli in ways that would allow information about individual features to be readily *encoded* by the amplitude of the BOLD signal. To do so, we employed the so-called *frequency tagging* technique (Regan, 1989), which has been previously used in EEG recordings (Andersen, Hillyard, & Muller, 2008; Muller et al., 2006; Schoenfeld et al., 2007). In this technique, the presentation of each stimulus feature is modulated in time at a specific temporal frequency so that the evoked response associated with that feature could be extracted from the harmonic response at that frequency. Accordingly, the response evoked by multiple features can be readily teased apart by tagging each feature with its own unique frequency.

We implemented this strategy in a stimulus that consisted of two superimposed transparent surfaces, each comprised of a field of dots with a distinct color–motion conjunction that changed smoothly with time. The frequency tagging was performed by making both the direction of motion and the color of the dots change periodically, and with distinct frequencies. We used this experimental setting to determine the effect of feature- and object-based attention on BOLD signals throughout visual cortex. By analyzing the amplitude of the BOLD response at each of the four designated frequencies, we found that attention modulated the response to the attended feature as well as the task-irrelevant feature associated with the same surface. This effect, which was present in multiple visual areas including V1, demonstrates that object-based attention modulates feature-specific representations across the visual cortex.

METHODS

Participants

4 male and 4 female subjects aged 20 to 28 gave written consent in accord with the human subjects' protocol at the University of Washington to participate in this study. They all had normal or corrected-to normal vision and 6 of them were naïve to the purpose of the experiment. Subjects participated in two separate experiments: (1) an fMRI experiment that consisted of one retinotopic mapping session followed by two 2-hr functional scanning sessions, and (2) a psychophysical experiment that consisted of two 1-hr behavioral sessions. Four of the eight subjects participated in both experiments. For both fMRI and psychophysical experiments, subjects completed 1-2 hours of training to ensure that they were familiar with the task.

Stimulus

The stimulus used in both the fMRI and behavioral experiments consisted of two superimposed fields of dots. Each dot field consisted of 101 dots per frame (frame rate = 60 Hz) of the same color that moved coherently in a specific direction at a speed of 6 deg/sec. The two dot fields had distinct color-motion conjunction and appeared as two surfaces moving transparently across one another (Fig. 1A and Supplementary Movie 1). To remove a potential depth cue, the depth order of each dot (which dots occludes the other dots) was randomized. The two fields were rendered on a black background within an annulus with an inner diameter of 3° and an outer diameter of 16° of visual angle.

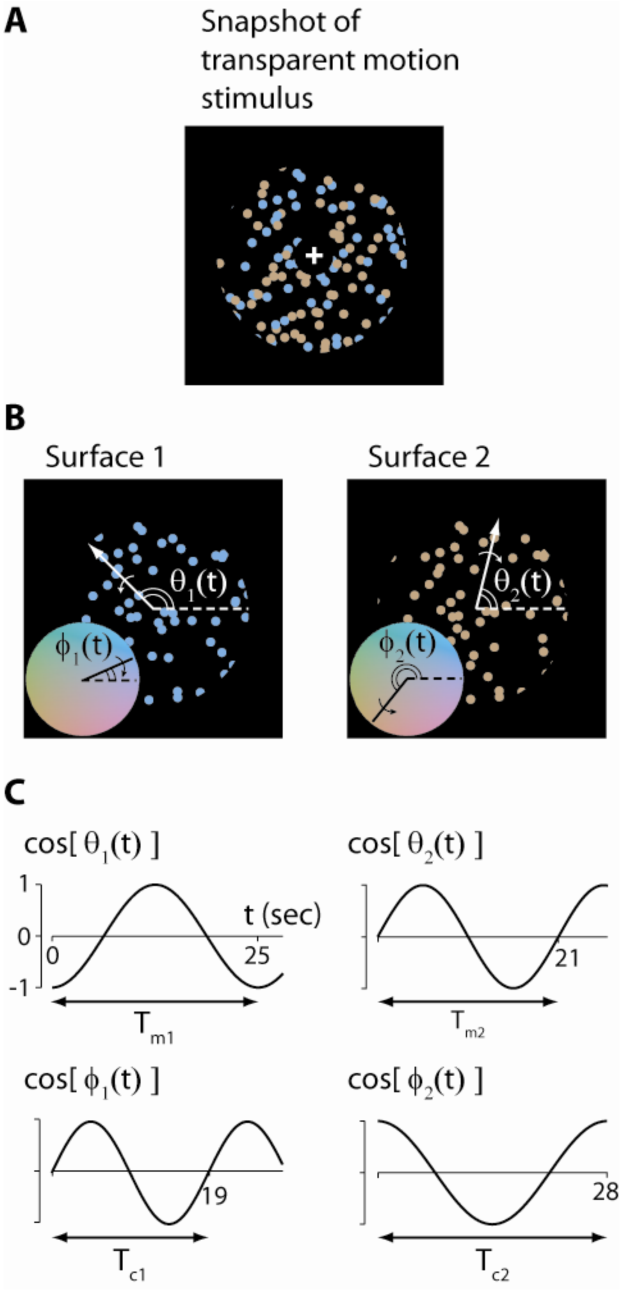

Fig. 1 *A*, The stimulus: two fields of superimposed dots confined to an annulus with a fixation cross at the center. *B*, the two surfaces have been separated in order to better illustrate the dynamic components of the two dot fields. Over the course of the stimulus movie, the direction of motion of the dots in surface 1 progresses counter-clockwise (white arrow). At a given point in time, *t*, following stimulus onset, the direction of motion of the dots were specified by $\theta_1(t)$, the angle from the dotted white line (rightward motion). The color of the dots in surface 1 progressed from blue to red to green—clockwise through the color space shown in the inset (black line). The

color of the dots was specified by $\phi_1(t)$, the angle from the dotted black line (blueish). The direction of motion of the dots in surface 2 progressed clockwise (white arrow), and were specified by $\theta_2(t)$, similar to surface 1. Likewise, the color of the dots in surface 2 progressed from red to blue to green—counter-clockwise through the color space shown in the inset—specified by $\phi_2(t)$. *C*, cosine functions of the four stimulus harmonics restricted to the first 28 seconds ($0<t<28$) of the stimulus movie. Each stimulus harmonic possessed a unique temporal period: 25 (T_{m1}) and 21 (T_{m2}) seconds for the motion of surface 1 & 2, respectively; 19 (T_{c1}) and 28 (T_{c2}) seconds for the color of surface 1 and 2, respectively.

At stimulus onset, the dots associated with one of the dot fields (hereafter, surface 1) appeared blue and moved leftward within the annulus. At the same time, the dots of the other dot field (hereafter, surface 2) appeared red and moved upward. During the presentation of the stimulus, the color and direction of motion of both dot fields changed slowly and cyclically.

Color specifications

The color of dots in each surface and at each time point was determined by a point in the CIE L*a*b* space. Changes in color with time were governed by slow movements of this point along a circular path through the CIE L*a*b* space (Fig. 1B,C). The CIE L*a*b* space was chosen for its perceptual uniformity in order to generate a color sequence that changed in chromaticity at a roughly constant rate. The two surfaces rotated along the same circular path but in opposite directions and with different temporal periods (Fig. 1B); surface 1 went from blue to red to green with a temporal period of 19.20 sec (T_{c1}), and surface 2 went from red to green to blue with a temporal period of 17.14 sec (T_{c1}). In the fMRI experiment, in which the stimulus was presented for a total of 8 min, surface 1 and 2 made 19 and 28 full rotations respectively. The circular path was defined mathematically as follows:

$$L^* = 100$$

$$a_i^*(t) = 42 \times cos\left(2\pi \frac{t}{T_{ci}} + \varphi_i + n\right) \tag{1}$$

$$b_i^*(t) = 42 \times sin\left(2\pi \frac{t}{T_{ci}} + \varphi_i + n\right)$$

where, i, indexes each surface (i= 1 or 2), t, represents elapsed time in seconds, and ϕ_i is a phase parameter which determines the surface's initial hue, and n corresponds to the brief dispersions that were added during a color event (see *Color events*). The constant 42 (amplitude of a* and b*) was chosen to keep the colors within the dynamic range of our projector. After specifying the L*a*b* values, we used standard CIE *XYZ* coordinate transformations to compute the corresponding RGB values to drive the calibrated projector.

To ensure that various colors were perceived as isoluminant, we adjusted the scaling of the RGB values using the so-called "minimum motion procedure" (Anstis & Cavanagh, 1983). Subjects viewed a stimulus that was made of two superimposed radial gratings, a chromatic test grating and a luminance grating. The two gratings were modulated sinusoidally and were in quadrature. The test grating was created from two phosphors modulating in anti-phase with a temporal period of 333 ms (20 frames) and one phosphor held at a constant intermediate intensity. The luminance grating was made of in-phase space-time modulations of the three phosphors at a Michelson contrast of 0.08. This stimulus is typically perceived as having clockwise or counter-clockwise apparent rotational motion unless the two modulating phosphors in the test grating are isoluminant.

We determined this point of isoluminance by a staircase procedure in which subject adjusted the ratio of the two test phosphors so as to null the apparent motion. On each trial, subjects fixated a central spot and used two buttons to reduce or increase the intensity of one of the two modulating phosphors of the chromatic grating (test phosphor) by a fixed amount (i.e., step size) to reverse the perceived direction of the apparent motion. After each reversal, the step size was halved and the procedure was repeated until subjects reported that they no longer perceived a clear apparent motion. We quantified this point of isoluminance by the ratio of the intensity of the test phosphor to the other modulating phosphor. To estimate the points of isoluminance across the whole RGB space, we repeated the staircase procedure for three different pairs of modulating phosphors (R-G, B-G and B-R) and estimate the three corresponding ratios of phosphor intensities (R/G, B/G, B/R).

To account for changes in isoluminance as a function of eccentricity, we measured RGB scale factors at three concentric non-overlapping annuli (2.167 deg width), spanning the spatial extent of our stimulus (1.5–8 deg in radius). For each subject, we made 3 independent measures of the 3 ratios (R/G, B/G, B/R) at each of the 3 eccentricities (total of 27 measurements), and used the corresponding averages as our estimate of the isoluminance ratios. The RGB scale factors did not change appreciably with eccentricity; nonetheless, we used the scales factors measured for the 3 eccentricities to derive a first order (linear) estimate of the RGB scale factor as a function of eccentricity, which we used to adjust the color of dots at different eccentricities. Each dot's color was further jittered by a small amount of luminance noise to counteract any small hue-dependent luminance bias that our isoluminance measurements failed to account for. The luminance noise consisted of white noise (std of 5.5 cd/m^2 low-passed filtered by a Gaussian in the frequency domain with std of 0.5 cycles/sec). Finally, we used the gamma curves for each phosphor to ensue the color outputs had the desired intensity.

Motion specifications

For each surface, the direction of motion was specified by the coherent translation of individual dots in a specific direction between successive monitor frames. The direction of motion of the two surfaces changed smoothly around the clock with temporal periods of 25.25 (T_{M1}) and 21.82 sec (T_{M2}), which for an 8-min fMRI session, correspond to 22 and 25 full rotations respectively. The dynamics of the direction of motion in the two surfaces can be mathematically formulated as follows:

(2) $$\theta_i(t) = 2\pi \times \frac{t}{T_{mi}} + \varphi_i + n$$

where, θ corresponds to the direction of motion in radians, t corresponds to the elapsed time in seconds, i indexes each surface, and φ is a phase parameter which determines the surface's initial motion direction, and n corresponds to the dispersion added during a motion event (see *Motion events* bellow).

The color and direction of motion of the dots were subject to a number of constraints. We use the terms "birth", "death", "age" and "lifetime" to refer to the beginning, the end, the number of elapsed frames since birth, and the duration (in frames) of coherent translation of a dot. At stimulus onset (t=0) each dot was assigned a random age from 0 to 11. The initial color and motion of each dot was specified by setting t=0 (Equation 1 and 2). After each monitor refresh, the age of every dot was incremented by 1. Each dot maintained its color, its luminance and its direction for the duration of its lifetime, which was fixed to 12 frames (200 ms). If a dot moved outside the annulus during its lifetime, then its position was wrapped around to the other side of the annulus. The death of each dot (age=12) led to the birth of a new dot (age=0) at a random position within the annulus. The color and direction of motion of each newly born dot were adjusted based on Equations 1 and 2 at the new t (12 frames past the previous birth). Because the dots' ages at stimulus onset were assigned randomly, at any given moment in time, each surface maintained a small dispersion around its average color and direction of motion.

Stimulus events

To provide a task for the subjects, each surface was subject to brief perturbations in color and direction, which we refer to as "color events" and "motion events" respectively. Each event type (color and motion) in each of the two surfaces occurred multiple times with an average frequency of once per 7 seconds. Each event lasted one second, and was followed by an absolute refractory period of 1 sec.

Color events

A color event was characterized by a transient increase in the variance of the color in a surface, which lasted a total of 1 sec. To modulate the variance, we added noise to the color of individual dots that were born during the event. The additional noise was controlled by a random variable, n, that was added to the phase of a^* and b^* (Equation 1). In the absence of a color event, the value of n was set to zero (i.e., no additional variance). During a color event, n went from 0 to n_{max} for the first 0.5 sec of the event, and then went back to 0 for the second half. For each dot, n_{max} was specified by a random draw between -2π/3 and 2π/3 radians.

Motion events

Similar to the color event, when assigning a direction of motion to a dot born during a motion event, a random variable, n, (Equation 2) was added to the phase of the direction of motion. The value of n was set to zero when no motion event were present. During a motion event, the value of n went from 0 to n_{max} for the first 0.5 sec of the event, and then went back to 0 for the second half. For each dot, n_{max} was specified by a random draw between -2π/3 and 2π/3 radians.

fMRI scanning sequence

The stimulus was identical across scans. Prior to a scan, subjects were cued to track one of the four surface features, or perform a demanding task at fixation. Subjects where scanned over two separate scanning sessions, held on separate days. The sequence of tasks in a typical scanning day was as follows: (1) the fixation task; (2) track the motion of surface 1; (3) track the motion of surface 2; (4) repeat the fixation task; (5) track the motion of surface 2; (6) track the motion of surface 1. For half of the subjects, the direction of motion was probed on day 1 and color on day 2; for the other half, this sequence was reversed.

Behavioral task: fMRI

For each scan, observers were cued to track a single surface and to detect events within a single feature, either the motion or color of that surface. We refer to this event as the target event and the other three event types as distractor events. For example, when instructed to detect events in the color of surface 1, the color events in surface 1 were the target events, and the motion events in surface 1 as well as both color and motion events in surface 2 were distractor events. Observers were instructed to press a response button immediately after detecting every target event, while ignoring all distractor events.

Behavioral analysis: fMRI

Responses were divided into (1) hits, (2) false alarms, and (3) selection errors (misses were separately tallied). A response was classified as a hit if the subject pressed the response button within a two second window after the onset of the target event. For each scan, the hit rate was computed by dividing the number of hits by the total number of target events. Target events that were not followed by a button press were referred to as misses. A response was classified as a false alarm if no event (target or distractor) preceded the response within a two second window. Finally, a response was classified as a distractor response if the two seconds preceding the response contained any distractor event but no target-event.

Psychophysics experiment

On each trial subjects were presented with a 5-sec stimulus, similar to the one we used in the fMRI experiment. A random initial color (ϕ_1) and motion direction (θ_1) was selected for the first surface on every trial. The initial color and motion direction of the second surface was shifted by 90 deg in color- and motion-space from the first surface; i.e., $\phi_2 = \phi_1 + \pi/2$ (Equation 1), and $\theta_2 = \theta_1 + \pi/2$ (Equations 2).

There were two single-task conditions and one dual-task condition. Subjects were either cued to track the color, the motion, or both color and motion within one of the surfaces. At 300 ms post stimulus onset, the color and/or the motion of one surface was cued. To cue motion, the speed of the cued surface increased; to cue color, the luminance of the cued surface increased. In both cases the intensity change was stepwise, lasting 300 ms.

Color and motion events occurred with 50% probability within each feature. The onset of an event occurred randomly, and with equal probability, between 1 and 4 seconds from stimulus onset. To prevent subjects from using a switching strategy in the dual-task condition, on trials when both a color and motion event occurred within the same surface, they were constrained to occur simultaneously. Thus, on 25% of the trials the motion and color events co-occurred. Following stimulus offset, subjects reported (via the key press) whether or not there was any target event in the stimulus. The yes/no responses for each task were mapped to separate keys, one set for each hand. The response order was counter balanced across subjects in the dual-task condition.

Subjects performed each condition in blocks of 30 trials. The order of the blocks was counterbalanced between subjects. Subjects practiced the task over one or two 1-hour sessions. After practice, each subject ran four blocks of 30 trials for each condition, for a total of 120 trials per condition.

Retinotopic mapping procedure

Retinotopic mapping was obtain in a single one hour session using a flickering checkerboard restricted to a rotating wedge, an expanding annulus, and an alternating pair of wedges covering the vertical and horizontal meridian [stimulus flicker 6 Hz, wedges subtended 40° of polar angle (Engel et al., 1994; Sereno et al., 1995)]. With this procedure, V1, V2v, V2d, V3v, V3d, and hV4 were drawn by hand on the inflated representation of the cortical surface using BrainVoyager QX (version 1.9.10 Brain Innovation, Maastricht, The Netherlands). Ventral and dorsal areas were collapsed together for the analysis. Ventral area hV4 was defined to include an entire hemifield representation (Wandell et al., 2005). A functional localizer was used during each experimental session to define MT+.

fMRI data acquisition and analysis

MRI scanning was performed on a Phillips Achieva 3-Tesla scanner, located at the University of Washington Magnetic Resonance Research Laboratory, equipped with an eight-channel head coil. Anatomical T1-weighted images were acquired at 1 x 1 x 1 mm resolution. Whole-brain, 32 transverse slice, functional images were acquired at 3.438 x 3.438 x 3.5 mm resolution (repetition time, 2000 ms; echo time, 30 ms; flip angle, 76°; scan resolution, 64 x 64; field of view, 220 mm; slice thickness, 3.5 mm; no gap).

Each scan was motion corrected using BrainVoyager QX. Experimental scans were co-registered to the anatomical retinotopy scans. Custom software, written by John Serences, was used to extract the time-courses of voxels corresponding to each pre-defined visual area and account for the mismatch in resolution between functional and anatomical images.

Region of interest selection

Localizer scans were run at the beginning and end of each experimental session. A general linear model (GLM) was used to find voxels that responded strongly to the region of visual space corresponding to the extent of the stimulus. Regressors were created in the GLM by convolving a gamma function with the boxcar stimulus protocol. The functional localizer consisted of 20 second blocks of fully coherent moving achromatic dots (randomly reassigned one of eight possible direction of motion every second, 200 ms limited life time), static dots (redraw in a random configuration every second), and a blank screen. All other properties of the localizer were set to match the experimental configuration, including the dimensions of the stimulus aperture, as well as dot size, speed, density, etc. We selected voxels in V1, V2, V3, and hV4 that responded more strongly to the motion than to the blank condition. We defined MT+ as a contiguous patch of medial temporal cortex that responded more to the motion condition than the static condition ($p < 0.05$, Bonferroni corrected for multiple comparison).

Frequency analysis

We quantified the periodicity of hemodynamic response at the frequencies associated with the color and direction of motion of the two surfaces from the amplitude of the corresponding harmonics in the Fourier spectrum. We measured the Fourier spectrum of responses in different visual areas by applying MATLAB's fast Fourier transform (FFT) to the average time course of the BOLD signal in those areas.

RESULTS

Subjects were scanned while viewing a stimulus consisting of two superimposed surfaces composed of dot fields with unique color and motion conjunction (Fig. 1A). The direction of motion and color of each surface slowly changed in a periodic fashion with a unique temporal period for each feature (Fig 1B,C and Supplementary Movie 1). Prior to a scan, subjects were cued to track the motion or color of one of the two surfaces and performed an ongoing task in which they were instructed to respond with a button press every time they detected a target event in the cued surface feature (while ignoring distractor events). Events were defined as brief dispersions in the motion or color coherence (see *Methods: stimulus events*).

Behavioral results during fMRI data acquisition

All six subjects were able to track the cued surface-feature in order to respond to target events and ignore distractor events. Table 1 shows the proportion of button presses following target events (target response), distractor events (distractor response), and no events (false alarms) for each subject. The majority of button presses were associated with a target event (0.91±0.04, mean ± std). There were relatively fewer distractor responses (0.07±0.04), and very few false alarms (0.02±0.02). The high ratio of target responses to distractor responses suggests that subjects were able to track the cued surface-feature as it modulated in feature space. We intentionally made the magnitude of the event transients small, so as to reduce distractor interference. This resulted in a difficult detection task, so even though most responses were target responses, the miss rate was still high (0.44±0.06).

Subject	1	2	3	4	5	6
Target response	0.86	0.96	0.91	0.90	0.89	0.96
Distractor response	0.13	0.04	0.07	0.09	0.05	0.03
False alarm	0.01	0.00	0.02	0.01	0.06	0.01

Table 1. Proportion of button responses for the detection task in the scanner for each subject. Target response: percentage of responses following a target event; distractor response: percentage of responses following a distractor event; False alarm: percentage of responses following no target or distractor event

The harmonic hemodynamic response

Previous work has shown that individual voxels in different visual areas could exhibit weak but reliable selectivity for stimulus features including color and direction of motion (Fig. 2A) (Brouwer & Heeger, 2009; Kamitani & Tong, 2006). Consequently, by changing the color and direction of motion of the stimulus in a circular fashion, we should be able to modulate the response of color- and direction-selective voxels in a periodic fashion (Fig. 2B). Because the BOLD signal is sluggish we used relatively low frequencies (long periods) over which to modulate the stimulus features: 25 and 22 cycles/scan (8 minutes/scan) for the direction of motion, and 19 and 28 cycles/scan for the color of surface 1 and 2 respectively. We hypothesized that attention modulates the gain of feature-selective neurons (Martinez-Trujillo & Treue, 2004), leading to a change in the amplitude of the corresponding harmonics in the hemodynamic response (Boynton, 2005a). We therefore analyzed the frequency spectrum from the BOLD time course in different visual areas and compared the amplitude of the harmonic responses associated with color and direction of motion of the two surfaces across attention conditions (Fig. 2C).

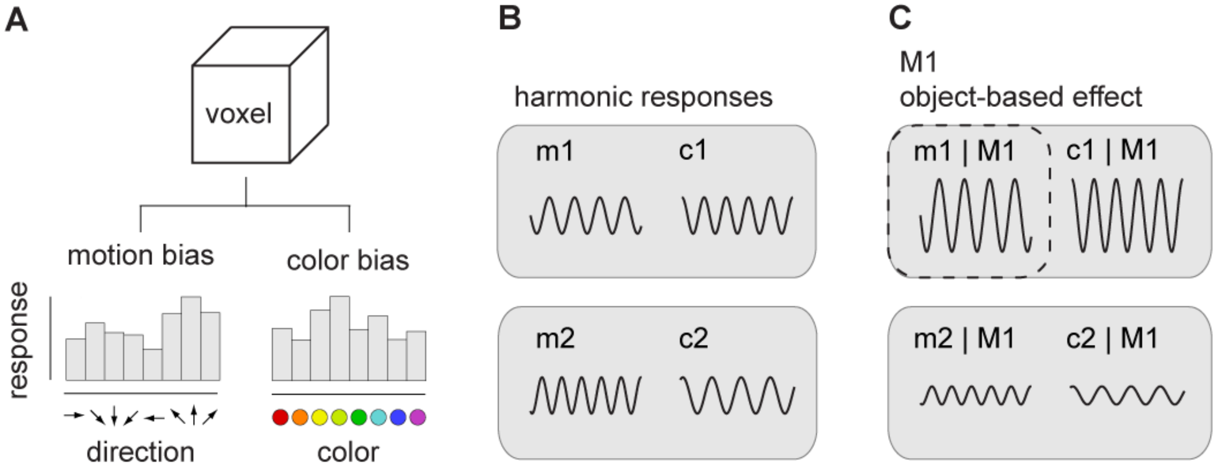

Fig. 2 *A*, A hypothetical voxel's response to different directions of motion (lower left of panel *A*) and different hues (lower right of panel *A*). The response is illustrated as exhibiting a small bias for upward motion and greenish hues. Accordingly, the voxel would exhibit a harmonic response to stimuli whose direction of motion or color is modulated periodically. *B*, Illustrates a temporal segment of the independent harmonic response to each stimulus feature. A motion and color component for surface 1 (*m1* & *m2*, respectively), and a motion and color component for sirface 2 (*m2* & *c2*, respectively). *C*, When one surface feature is attended (e.g. the motion of surface 1: *M1)*, a gain model of feature-based attention predicts an enhanced response to that feature (illustrated here as an increase in $A_{m1}|M1$). A gain model of object-based attention makes the additional prediction that the amplitude of the harmonic response to the task-irrelevant feature (the color of surface 1: $A_{c1}|M1$) is also modulated by attention.

To facilitate the presentation of results, we developed nomenclature to refer to the different attention conditions and different response harmonics: A_{m1} and A_{m2} are the amplitudes of the harmonic responses associated with the motion feature in surface 1 and 2 respectively; A_{c1} and A_{c2} are the amplitudes of the harmonic responses associated with the color feature in surface 1 and 2 respectively; *M1* and *M2* refer to conditions in which subjects tracked the motion of surface 1 and 2 respectively; *C1* and *C2* refer to conditions in which subjects tracked the color of surface 1 and 2 respectively.

We collapsed the responses across all voxels within each visual area and then averaged across subjects to generate an average amplitude spectrum for each attention condition. The between-subjects averaged amplitude spectrum for V1 is shown in Fig. 3 for each attention condition. Fig. 3A shows the amplitude spectrum when the motion of surface 1 was tracked (*M1*). The amplitudes labeled $A_{m1}|M1$ and $A_{m2}|M1$, at 22 and 25 cycles/scan, refer to motion harmonics to surface 1 and 2 respectively, and $A_{c1}|M1$ and $A_{c2}|M1$, at 28 and 19 cycles/scan, refer to the corresponding color harmonics. In this condition, the $\mathrm{A_{m1}|M1}$ was stronger than baseline (harmonic amplitude at the non-stimulus frequencies) suggesting that attention to *M1* caused an enhancement of response to the corresponding feature. However, this effect was not due to a non-specific enhancement of direction-selective mechanisms in V1 because the same response modulation was not present for the motion in surface 2 ($A_{m2}|M1$). The lack of response enhancement to the other overlapping surface also rules out the possibility that the modulation of $A_{m1}|M1$ was due to spatial attention. The same selective enhancement of response was evident in the condition in which subjects tracked the motion of the surface 2 (*M2*): $A_{m2}|M2$ but not $A_{m1}|M2$ was stronger than baseline (Fig. 3C). These results indicate that, in V1, the effect of feature-based attention to the direction of motion was surface specific.

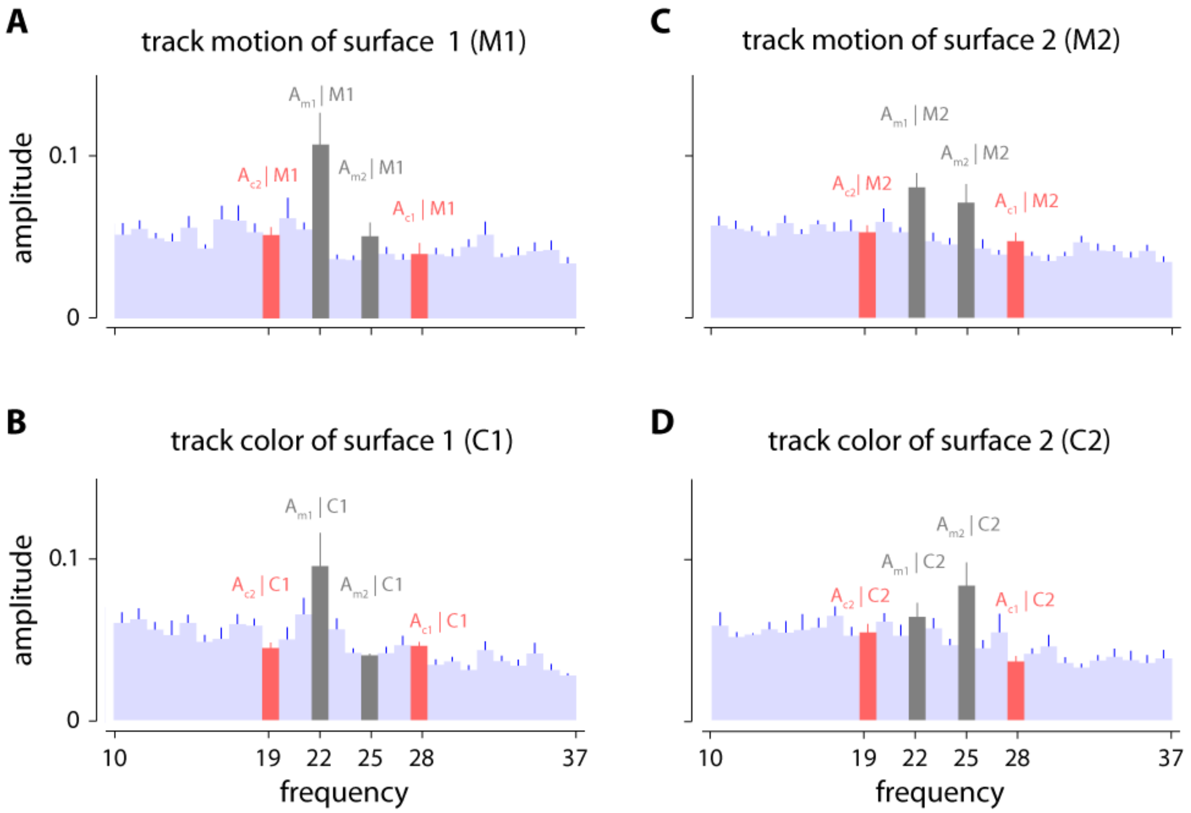

Fig. 3 Between subject average amplitude spectrum for area V1. The error bars encompass ±1 standard error of the mean across the 6 subjects. A subsection of the full amplitude spectrum is shown (10–37 cycles per scan) to focus on the amplitudes of the four stimulus harmonics: A_{m1} & A_{m2}, highlighted in gray at 22 & 25 cycles per scan, respectively, and A_{c2}, & A_{c1}, highlighted in red at 19& 28 cycles per scan. The remainder of the frequency spectrum (in blue) corresponds to non-stimulus harmonics (noise response). *A, B*, The average amplitude spectrums for the *M1* and *C1* conditions, respectively. *C, D*, the average amplitude spectrums for the *M2* and *C2* conditions, respectively.

In contrast, when subjects attended the color of either surface, the color harmonic responses in V1 (A_{c1} | *C1* and A_{c2} | *C2*) were not different from baseline (Fig. 3B,D). Surprisingly however, attention to color of a given surface led to an enhancement of responses associated with the motion of that surface. In particular, in the *C1* and *C2* conditions respectively, A_{m1} | *C1* (Fig. 3B) and A_{m2} | *C2* (Fig. 3D) were significantly stronger than the noise harmonics. In fact the overall pattern of amplitude responses across the four stimulus frequencies looks similar regardless of whether the subjects tracked the motion or color of a surface (Fig. 3, A versus B, and C versus D). These results suggest that the observed modulations in V1 BOLD responses are associated with surface specific attentional selection for the direction of motion.

A similar pattern of responses was observed in visual areas V2 and V3 (supplemental Fig. 1&2). Area V4 followed the same trend but was less reliable than the earlier visual areas (supplemental Fig. 3). Area MT+ did not produce a reliable harmonic response to our stimulus (supplemental Fig. 4).

In a separate experiment, we attempted to measure a baseline response to the stimulus while subject's attention was diverted by a fixation task. Under this attentional condition the stimulus failed to drive voxels at the stimulus harmonics across all visual area (data not shown). Attention to the stimulus was therefore required to drive the harmonic response above the noise.

Feature-based attention

The selective enhancement of A_{m1} and A_{m2} in $M1$ and $M2$ conditions respectively corresponds to the effects of feature-based attention for the direction of motion (Fig. 3A,C). We developed a feature-based attention index (FI) to quantify the magnitude of feature-based attention for both motion and color in different visual areas. The index measures the relative change in the amplitude of the harmonic response to a given feature when it is attended versus when it is unattended. For example, FI_{m1} is the normalized difference between the amplitude of the harmonic response to the motion of surface 1 (A_{m1}) under two different attention conditions, $M1$ and $M2$.

$$FI_{m1} = \frac{A_{m1}|M1 - A_{m1}|M2}{A_{m1}|M1 + A_{m1}|M2} \tag{4}$$

This formulation of the FI provides a simple metric for feature attention that ranges between -1 and 1, with zero corresponding to no attentional effect and 1 and -1 corresponding to strong enhancement and suppression of responses. Figures 4A,B show FI for direction of motion (FI_{m1} and FI_{m2} for surface 1 and 2 respectively) for each visual area. For both surfaces, the FI was positive in V1, V2, V3 and V4, but the effect was weak and only significant in areas V1 and V2 for one of the surfaces ($t(6)=3.25$ for V1 and 2.80 for V2, $p<0.05$). The effect was weakest in area MT+ where direction selectivity is strong (R. B. Tootell et al., 1995; Zeki et al., 1991).

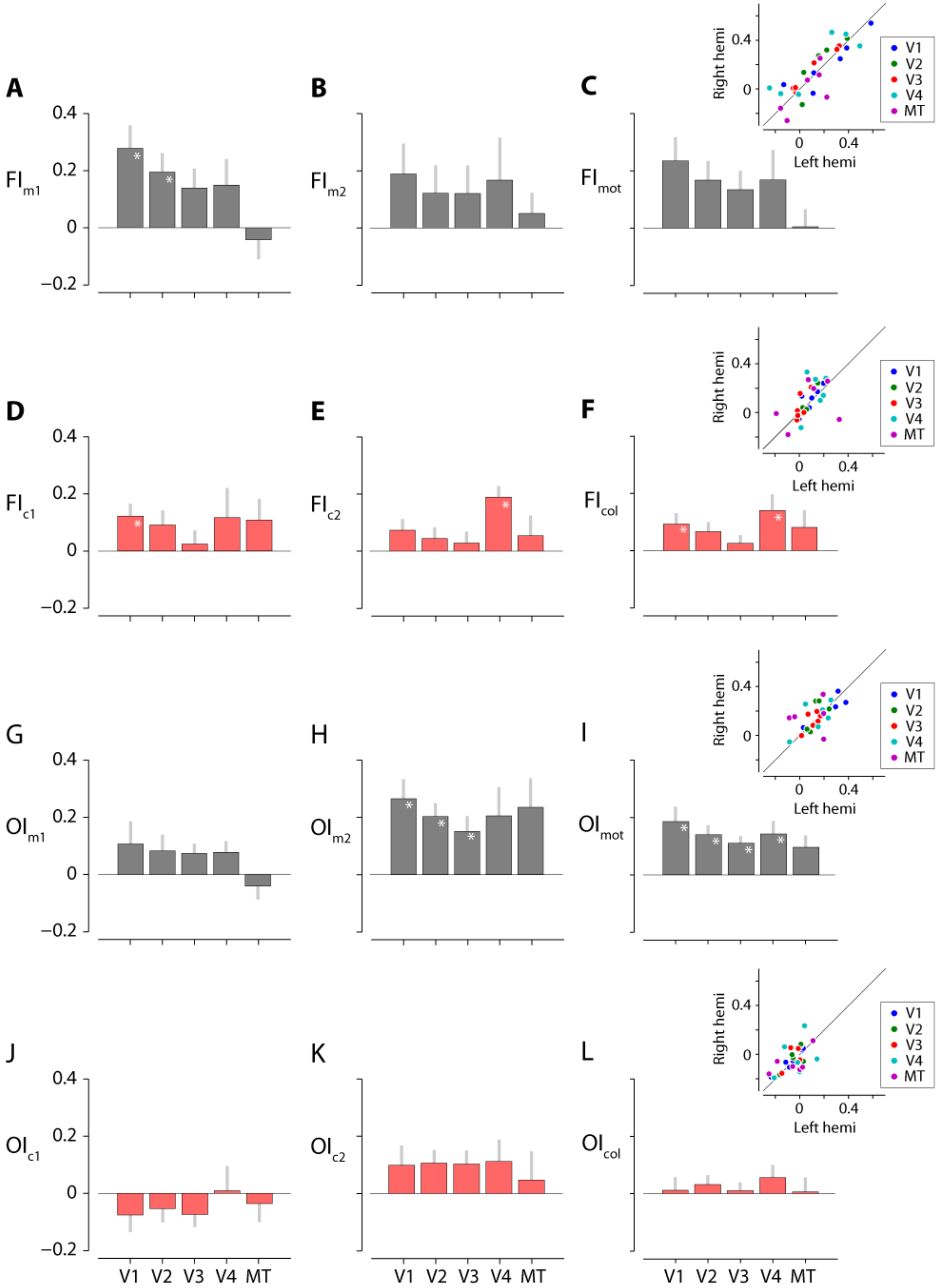
A
B
C
FI$_{m1}$
FI$_{m2}$
FI$_{mot}$
D
E
F
FI$_{c1}$
FI$_{c2}$
FI$_{col}$
G
H
I
OI$_{m1}$
OI$_{m2}$
OI$_{mot}$
J
K
L
OI$_{c1}$
OI$_{c2}$
OI$_{col}$
0.4
0.2
0
−0.2
V1
V2
V3
V4
MT
*
Right hemi
Left hemi
0
0.4
V1
V2
V3
V4
MT

Fig. 4 Feature-based and object-based attention indices across conditions and visual areas. Each panel shows the value of an attention-index averages across the 6 subjects for areas V1, V2, V3, V4, and MT+. *A-F,* Feature-attention indices. Feature-attention indices quantify the effect of attention to motion on motion harmonic responses (*A-C*), and the effect of attention to color on color harmonic responses (*D-F*). *A-C,* Feature-attention index for the direction of motion in surface 1 (FI_{m1}), the direction of motion in surface 2 (FI_{m1}), and the overall average of the two surfaces (FI_{mot}). *D-F,* Feature-attention index for the color in surface 1 (FI_{m1}), the color in surface 2 (FI_{m1}), and the overall average of the two surfaces (FI_{mot}). *G-L,* Object-attention indices. Object-attention indices quantify the effect of attention to color on motion harmonic responses (*G-I*), and the effect of attention to motion on color harmonic responses (*J-L*). *G-I,* Object-attention index for the direction of motion in surface 1 (OI_{m1}), the direction of motion in surface 2 (OI_{m1}), and the overall average of the two surfaces (OI_{mot}). *J-L,* Object-attention index for the color in surface 1 (OI_{m1}), the color in surface 2 (OI_{m1}), and the overall average of the two surfaces (OI_{mot}). In each row, the inset shows the overall average attention index computed from hemodynamic responses in the right hemisphere (ordinate) and left hemisphere (abscissa) for each subject and each visual area. The error bars show standard error of the mean across the 6 subjects. The grey asterisk denotes the attention-index values that were significantly different from zero ($p<0.05$, one-sample t-test).

In many cases, FI was not significant but showed a positive trend. To determine whether the weak positive trend in FI was reliable, we asked whether a similar trend was evident in both hemispheres. We first averaged FI_{m1} and FI_{m2} to compute a single index (FI_{mot}) for each visual area (Fig. 4C), and then compared this value between the two hemispheres for the 5 visual areas (V1, V2, V3, V4 and MT+) in the 6 subjects. This analysis (Fig. 4C, inset) showed that, across areas and subjects, FI_{mot} was significantly correlated between the two hemispheres ($r(30)=0.85$, $p<0.001$) suggesting that there was a weak but reliable feature-attention effect for motion.

We defined a similar feature index to quantify the effect of feature-based attention for color in different brain areas. Figure 4D,E show the FI for the color of the two surfaces (FI_{c1} for surface 1 and FI_{c2} for surface 2) for each visual area. The FI was positive across all visual areas for both surfaces, but the effect was only significant in V1 for surface 1 and V4 for surface 2 (t(6)=2.80 for V1, and 4.88 for V4, $p<0.05$).

We asked whether the effect of feature-based attention to color was reliable by comparing the FI values between hemispheres. To do so, we averaged FI_{c1} and FI_{c2} to compute a single index (FI_{col}) that quantified the effect of feature-based attention for color (Fig. 4F). Overall there was a positive effect of feature-based attention on the color response across visual areas, which reached significance in areas V1 and V4 (t(6)=2.36 & 1.34, $p<0.05$). We then compared FI_{col} in the left and right hemispheres across 5 visual areas (V1, V2, V3, V4 and MT+) in the 6 subjects (Figure 4F, inset), and found that FI_{col} was significantly correlated between the two hemispheres (r(30)=0.53, $p<0.01$), suggesting that the index is reliable.

Object-based spread of attention

The selective enhancement of the motion-driven components A_{m1} and A_{m2} in conditions where color was attended (*C1* and *C2*) suggests that attention to the color of each surface enhanced the motion-related signals of the attended surface (Figure 3B,D). To quantify this spread of attention to the task irrelevant features of a surface, we developed an object-based attention index (OI) to quantify the magnitude of object-based attention for both motion and color in different visual areas. The index measures the relative change in the amplitude of the harmonic response to a given feature when the other feature of that same surface is attended versus when the other surface is attended. For example, OI_{m1} is the normalized difference between A_{m1} under two different attention conditions, *C1* and *C2*.

(5) $$OI_{m1} = \frac{A_{m1}|C1 - A_{m1}|C2}{A_{m1}|C1 + A_{m1}|C2}$$

Like the FI, the OI ranges between -1 and 1. Figures 4G,H show OI to the direction of motion for each surface (OI_{m1} and OI_{m2} for surface 1 and 2 respectively) for each visual area. The OI_{m2} was positive in V1, V2, V3, and V4 for both surfaces, and was significantly greater than zero in V1, V2 and V3 (t(6)= 3.87, 4.39 and 2.83 respectively, $p<0.05$).

To summarize the effect of object-based attention on the motion response, we averaged OI_{m1} and OI_{m2} together to form OI_{mot} (Fig. 4I). OI_{mot} was positive in all five visual areas, and was significant in V1, V2, V3, and V4 (t(6)=3.43, 4.34, 4.66, and 2.82 repsectively, $p<0.05$). To test the reliability of this statistic we compared the OI_{mot} between hemispheres (Fig. 4I inset) across all five visual areas. The OI_{mot} was significantly positively correlated between hemispheres (r(30)=0.54, $p<0.01$).

Finally, we quantified the effect of object-based attention to the color harmonics under conditions in which the motion of each surface was cued. For these conditions, we found no consistent object-based attention: OI_{c1} and OI_{c2} were not significantly different from zero for any of the visual areas we tested. OI_{c1} was, on average, negative across visual areas (except V4) and OI_{c2} was positive (Fig 4J,K). The average effect (OI_{col}) across the two surfaces was positive but not significant (Fig. 4L), although the value of OI_{col} was significantly correlation between the two hemispheres (Fig. 4L inset; r(30)=0.58, $p<0.001$).

No dual-task cost when tracking two features within a surface

We conducted a separate psychophysical experiment outside the scanner to assess the behavioral consequences of our observed fMRI results. Subjects performed a yes-no detection task while viewing short 5-second segments of the same stimulus used in the fMRI experiment. In the single-task condition, subjects were either cued to track the motion or the color of one of the two surfaces. In the dual-task condition, subjects were cued to track both features within one of the two surfaces (see *Methods: Psychophysics experiment*).

We compared subjects' performance between the single- and dual task conditions to see whether dividing attention across features within a surface would result in a cost in behavioral performance. As evidenced by the scatter plot in Fig. 5, individual subjects' performance on the single- and dual-task conditions was comparable, with no significant difference across subjects (single-task minus dual-task: t(6)=0.62 p=0.56 for motion and t(6)=0.20, p=0.85 for color).

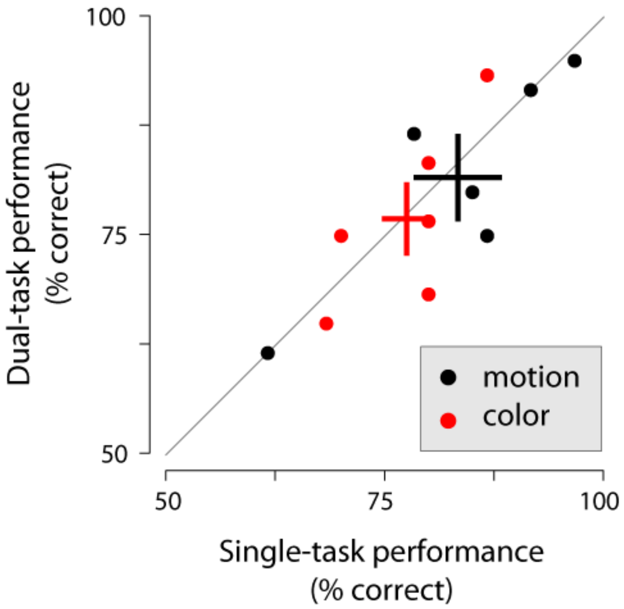

Fig. 8 Each subject's single-task performance (abscissa) is plotted against their dual-task performance (ordinate) for the motion task (black points) and color task (grey points). The identity line indicates equivalent single-task and dual-task performance. The black and gray crosses correspond to the average performance across the 6 subjects for the motion and color task respectively. The horizontal and vertical extent of each cross corresponds to the standard error of the mean for the single- and dual-task performance levels.

DISCUSSION

Both behavioral and physiological evidence has provided support for object-based attention. Theories of object-based attention posit that all features of a behaviorally-relevant object are selected. Behavioral studies support this claim showing that attention can be divided across multiple features within an object without a reduction in performance, whereas selecting features from different objects is much more difficult (Blaser et al., 2000; Bonnel & Prinzmetal, 1998; Duncan, 1984).

In an early neuroimaging study of object-based attention, O'Craven et al. (1999) presented subjects either a moving face superimposed on a stationary house, or a moving house superimposed on a stationary face. They found that when subjects attended one attribute of the moving face (either the face or the motion), the neural representation of both the motion sensitive area MT+ (MT/MST) and the fusiform face area were enhanced. Similarly, attending attributes of the moving house enhanced responses in MT/MST as well as the parahippocampal place area. This finding provides evidence that object-based attention enhances the response of brain areas representing the various attributes of a relevant object.

However, ROI-based analyses that rely on an averaged BOLD signal – like those used by O'Craven et al. (1999) – cannot specify whether object-based attention operates at the level of functional areas (e.g., area MT/MST) or at the finer level of feature-selective mechanisms within an area (e.g., direction selective neurons within MT). To overcome this limitation, several studies have exploited the inhomogeneities in sensory representations to decode feature-based attentional modulations from the pattern of hemodynamic responses (Brouwer & Heeger, 2009; Haynes & Rees, 2005; Kamitani & Tong, 2005, 2006; Serences & Boynton, 2007). For example, pattern classification has been used to predict which of two superimposed surfaces a subject attended (Kamitani & Tong, 2005). Pattern classification has also been used to classify object identity in higher visual areas (Kalanit Grill-Spector & Sayres, 2008; Haxby et al., 2001; O'Toole, Jiang, Abdi, & Haxby, 2005), and to decode the effects of selective attention on the representation of object-based information in extrastriate cortex (A. J.-W. Chen et al., 2012). Successful decoding provides evidence that selection of a feature could bias the pattern of responses across voxels towards the pattern produced by the attended feature in isolation (Boynton, 2005b). Here, we developed a frequency-tagging scheme that exploits these biases in order to assess the correlate of object-based attention at the level of feature-selective mechanisms within different visual areas.

Frequency-tagging with fMRI

Frequency tagging has been used to study the effect of selective attention in a number of EEG studies (Andersen et al., 2008; Muller et al., 2006; Schoenfeld et al., 2007; Toffanin, de Jong, Johnson, & Martens, 2009). To our knowledge, frequency tagging has not been used to study the effects of feature- or object-based attention with fMRI. But, frequency tagging is not new to fMRI; in traditional retinotopic mapping experiments, the spatial location of the stimulus is modulated in a circular fashion—both radially and tangentially—to infer the underlying retinotopic map from the phase of each voxel's response at the modulation frequency. The success of frequency tagging in retinotopic mapping is due to the large-scale cortical topography of receptive field locations in the visual cortex. This topography ensures that the profile of a voxel's response will oscillate at the frequency with which the retinal position of the stimulus is modulated.

The novel aspect of our work was to apply frequency tagging to analyze the responses to the features of color and direction of motion. We found that BOLD signals contained harmonic responses associated with the frequencies at which the color and direction of motion of the two surfaces were modulated. This finding suggests that some of the voxels contained an inhomogeneous distribution of feature-selective responses—for example, a weak preference for a particular color or direction of motion. The source of these inhomogemeties is not fully understood, but several possibilities have been proposed. Fine-scale anisotropies in the cortical organization of feature-selective neurons may give rise to the population response biases. Examples of such fine-scale anisotropies include the orientation-tuning columns and color-selective clusters in striate cortex (Hubel and Wiesel, 1974; Xiao et al., 2007), the columnar structure of direction-selective neurons in MT (Albright, 1984), and the color-selective clusters in extra-striate cortex (Conway et al., 2007). More recently, response biases at the voxel-based level was attributed to larger-scale biases in the topographic organization of the visual responses (Freeman, Brouwer, Heeger, & Merriam, 2011). In addition, anisotropies in the underlying representation of simple features across voxels could result in a measurable harmonic response across a population of voxels (Freeman et al., 2011; Mannion, McDonald, & Clifford, 2010; Op De Beeck, 2010).

Several constraints must be satisfied in order to combine frequency tagging with fMRI measurements. The first constraint is related to the sluggish nature of the hemodynamic response. Based on a typical model of the hemodynamic impulse response function (Boynton, Engel, Glover, & Heeger, 1996), short temporal periods, like those previously used to incorporate frequency-tagging in EEG recordings (Morgan, Hansen, & Hillyard, 1996), would not be extractable from the BOLD signal. Consequently, because the BOLD signal cannot capture rapid modulations of the underlying neural activity, we chose relatively long temporal periods (on the order of tens of seconds).

A second constraint in our design was our use of circular feature-spaces. To drive harmonic responses to a feature, it is important to be able to modulate that feature periodically. Color, direction motion, and orientation are all natural candidates since they can be readily represented and modulated in a circular fashion. It would be more difficult to use the frequency-tagging scheme to stimulus features that are not inherently circular such as spatial or temporal frequency. We chose color and direction of motion because they are natural features of dot fields, which can be easily superimposed.

Finally, frequency tagging implicitly assumes a linear relationship between the stimulus and the evoked responses. Evidence suggests that the hemodynamic response is to a first approximation linearly related to the average population response over time (Boynton et al., 1996). This assumption underlies a large body of fMRI research which uses an estimate of the hemodynamic impulse response function to predict the BOLD response to a time varying stimulus based on a general linear model (Heeger & Ress, 2002). However, the assumption of linearity is not without challenge (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001; Maier et al., 2008).

Specific to our stimulus, nonlinearities could also arise from neural populations tuned for specific color-motion conjunctions (Seymour et al., 2009). For example, if responses to the modulations of color and direction of motion interact multiplicatively, the harmonic responses would correspond to frequencies that are either higher or lower than those associated with the color and direction of motion in our stimulus. Consequently, our harmonic analysis was insensitive to the output of neurons that combine color and motion nonlinearly.

Neural correlates of feature- and object-based attention

We measured fMRI responses while subjects tracked either the color or the direction of motion of one of two overlapping surfaces (dot fields) segregated by their unique conjunctions of color and motion. We designed the stimulus such that the four features—two colors and two directions of motion—smoothly traversed a circular path through feature space with four unique temporal periods. This design enables us to use the corresponding harmonic responses to infer the effects of both feature- and object-based attention.

We found that the feature- and object-based effects were qualitatively similar (compare Fig. 3*A* to *B*, and *C* to *D*). For example, A_{m1} (the amplitude of the harmonic associated with the motion of surface 1), was modulated in both the M1 and C1 conditions (Fig. 3*A,B*). Our result supports the hypothesis that object-based attention modulates the sensory representation of all the features that comprise a task-relevant surface, which is consistent with previous fMRI (O'Craven et al., 1999) and electrophysiological measurements (Fallah, Stoner, & Reynolds, 2007;

Katzner, Busse, & Treue, 2009; Roelfsema, Lamme, & Spekreijse, 1998; Wannig, Rodríguez, & Freiwald, 2007). Our application frequency tagging to fMRI data suggests that the mechanisms of object-based attention—previously quantified at the level of average BOLD signal across visual cortical areas—might operate at a finer level of feature-selective mechanisms within those areas. Furthermore, our findings extend the effects of object-based attention to the hemodynamic response in area V1. Although feature-based attention effects have been reported in V1 (Saenz et al., 2002), object-based attention effects have only been reported in extrastriate visual areas including MT+ (Katzner et al., 2009; O'Craven, Rosen, Kwong, Treisman, & Savoy, 1997), the fusiform face area, and the parahippocampal place area (O'Craven et al., 1997). Our results extends previous work and suggests that object-based attention rely on feedback signals that target the representation of feature-specific mechanisms throughout the visual cortex and as early as V1.

The attention indices we used to quantify the effects of attention were based on comparing the hemodynamic responses to a feature under two different attentional states. This relative measure does not specify whether an increase in the attention index represents enhancement of responses to the attended feature or suppression of responses to the unattended feature, or a combination of both. Distinguishing between these possibilities requires an estimate of the baseline response to the stimulus. We attempted to measure a baseline response to our stimulus in a separate *fixation scan* in which subjects performed a demanding fixation task in order to draw their attention away from either surface. Interestingly, with attention directed to fixation, the peripheral stimulus failed to produce a reliable response at any of the stimulus harmonics; the amplitude at the stimulus frequencies was indistinguishable from the surrounding noise frequencies. Therefore, we were unable to determine the nature of attentional modulations when attention was oriented towards one of the two surfaces.

Signal-to-noise in our measurement

The average attention index for the cued feature and the task-irrelevant feature on the cued surface was positive in all 5 regions of interest. However, our effect size was small. Several factors could have contributed to a weak overall harmonic response. For example, it is thought that the superimposition of competing features within the receptive field of feature-selective neurons could reduce sensory-evoked responses (Desimone, 1998; Moran & Desimone, 1985; Reynolds, Chelazzi, & Desimone, 1999). Such suppressive effects are likely to degrade the signal-to-noise of our measurements. Another contributing factor may be the nonlinearities in feature-selective responses. Our frequency-tagging approach is only able to extract signals that are driven by a linear combination of responses to individual features. Therefore, the inherent nonlinearities in sensory representations and the potential interactions between the four features in our stimuli may have further reduced the sensitivity of our measurements to feature- and object-based attentional modulations. Finally, to combine frequency tagging with the fMRI signals, it was important to modulate the color and direction of motion of the two surfaces using relatively long temporal periods. Such slow modulations could adapt central feature-selective mechanisms (Boynton & Finney, 2003; Liu, Larsson, & Carrasco, 2007) and reduce the amplitude to the associated hemodynamic responses (K Grill-Spector & Malach, 2001).

The attention index for the motion harmonic was weaker in area MT+ than earlier visual areas; a surprising result given that large effects of attention have been previously reported in area MT+ (O'Craven et al., 1997; Saenz et al., 2002; Serences & Boynton, 2007; R. B. H. Tootell et al., 1998). This unexpected result might be related to a suppressive interaction between the two surfaces as shown by electrophysiological recordings in the macaque monkey (Treue, Hol, & Rauber, 2000), which is consistent with the weak harmonic responses out stimuli evoked in area MT+ (Supplemental Fig. 4).

Across visual areas, the attention index was smaller for the color harmonics than for the motion harmonics. We do not know the factors that contributed to a weaker color response. One possibility might be that the distribution of color selective neurons might be more homogeneous within a voxel than the distribution of motion selective cells, effectively leading to a smaller color harmonic. But this seems unlikely since color has been successfully classified in a number of studies using fMRI (Brouwer & Heeger, 2009; Kamitani & Tong, 2006; Seymour, Clifford, Logothetis, & Bartels, 2009). Another possibility might be related to the constraints of creating isoluminant stimuli, which limited the range of intensities we were able to use to modulate the red and green channels (to balance the luminance of the weaker blue channel). Finally, it is possible that direction of motion is inherently more effective that color in segregating transparent surfaces. If so, it is possible that the even when subjects were asked to track the color of a surface, they still implicitly used the motion cue to improve their ability to segregate the two surfaces.

Perceptual consequences of object-based attention

Object-based attention allows subjects to divide attention to multiple features of an object with no additional cost. Our fMRI experiment showed that attention to one feature of a surface modulated the neural response of that feature as well as the other feature of the same surface, but our behavioral paradigm did not directly test the consequences of divided attention. We therefore used our stimuli in an additional psychophysical experiment to establish a more direct link between our fMRI experiments and the mechanisms to object-based attention. Following previous work (Blaser et al., 2000; Bonnel & Prinzmetal, 1998; Duncan, 1984), we compared subjects' performance across two conditions. In the single-feature condition, subjects were required to detect changes in one of the surface features, and in the dual-feature condition, they detected changes in both features of the surface. We found that subjects' performance was comparable between the two conditions (Fig. 5) suggesting that the capacity to attend multiple features of an object might be mediated by the simultaneous enhancement of sensory responses to the features of that object.

SUPPLEMENTAL FIGURES

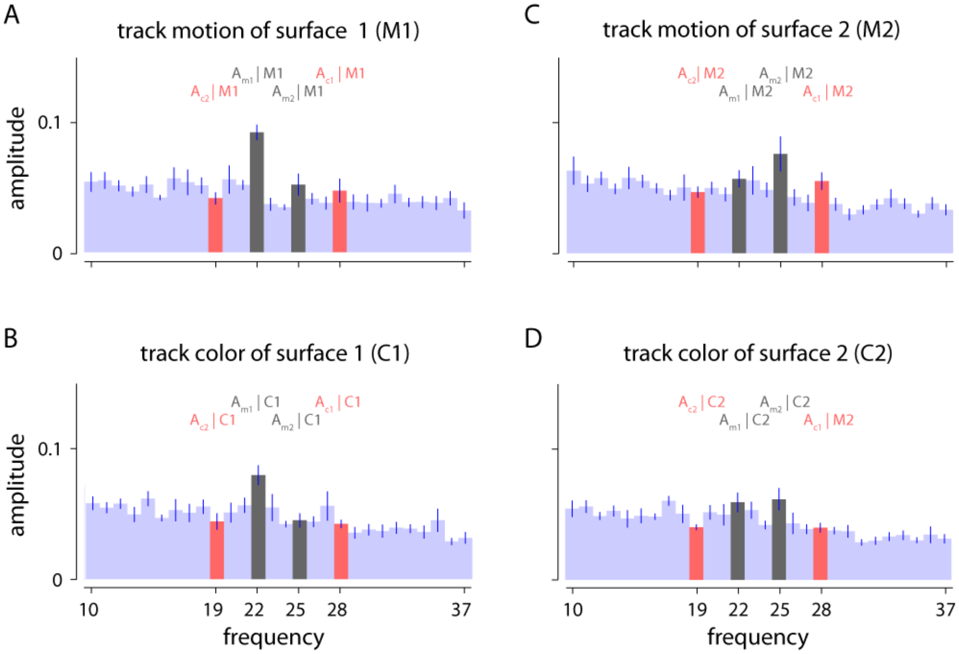

Fig. 1 Between subject average amplitude spectrum for area V2.

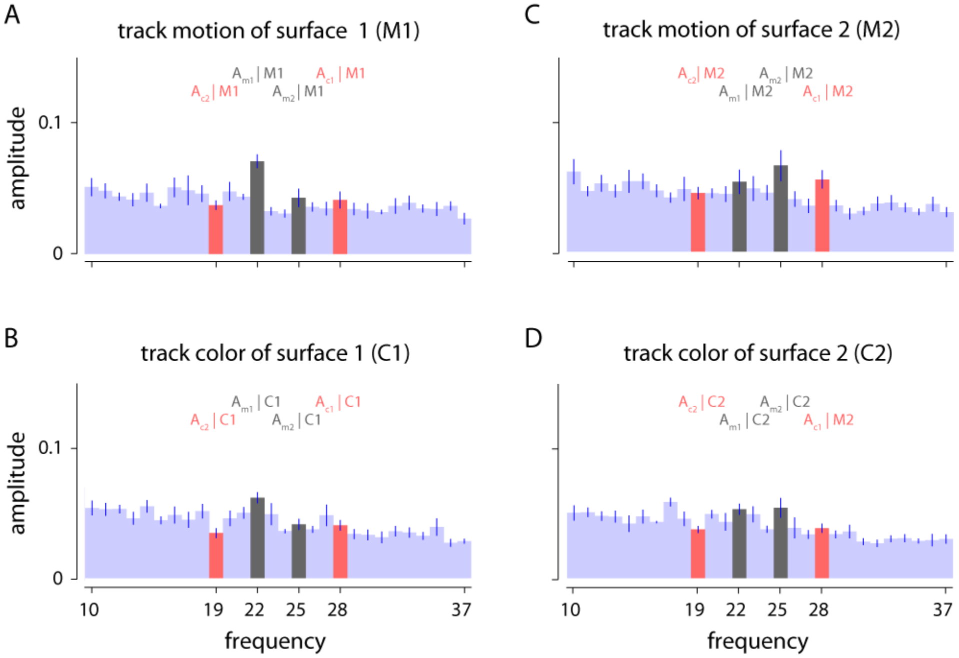

Fig. 2 Between subject average amplitude spectrum for area V3.

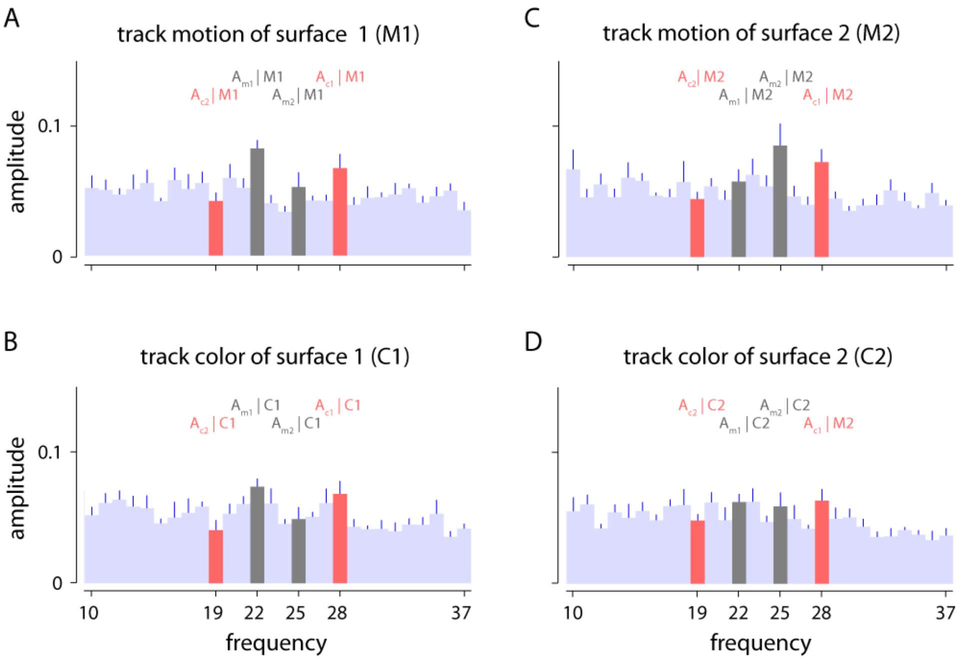

Fig. 3 Between subject average amplitude spectrum for area V4.

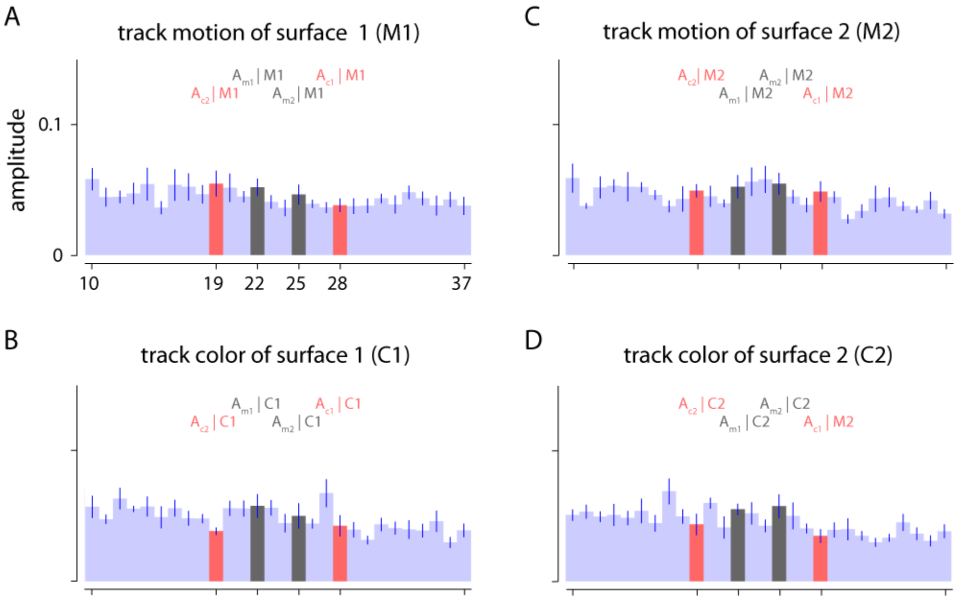

Fig. 4 Between subject average amplitude spectrum for area MT+.

DIVIDING ATTENTION BETWEEN TWO TRANSPARENT MOTION SURFACES

In collaboration with:
John Palmer & Geoff Boynton
Submitted to the Journal of Vision: May 4th, 2012
(under review)

In object-based attention, it is easier to divide attention between features within a single object than between features across objects. In this study we test the prediction of several capacity models in order to best characterize the cost to dividing attention between objects. Here we studied behavioral performance on a divided attention task in which subjects attended to the motion and luminance of overlapping random dot kinemategrams, specifically red upward moving dots superimposed with green downward moving dots. Subjects were required to detect brief changes (transients) in the motion or luminance within the same surface or across different surfaces.

There were two primary results. First, the dual-task deficit was large when attention was divided across two surfaces and near zero when attention was divided within a surface. This is consistent with limited-capacity processing across surfaces and unlimited-capacity processing within a surface—a pattern predicted by established theories of object-based attention. Second and unexpectedly, there was evidence of crosstalk between features: when cued to monitor transients on one surface, response rates were inflated by the presence of a transient on the other surface. Such crosstalk is a failure of selective attention between surfaces.

INTRODUCTION

More sensory information is available to the visual system than can be effectively processed. Due to these limits in processing, information competes for memory encoding, perceptual decisions, and motor responses. Visual attention helps resolve this competition by selecting relevant information on the basis of spatial location, feature, or object, which biases the sensory processing in favor of the behaviorally relevant information (Posner, 1980; Desimone & Duncan, 1995; Desimone, 1998; Kanwisher & Wojciulik, 2000). Object-based attention is hypothesized to select all the features of a behaviorally relevant object, serving to improve the encoding of its component features relative to an unattended object (Duncan, 1984; Kahneman, Treisman, & Gibbs, 1992; Treisman, 1998; Valdes-Sosa, Cobo, & Pinilla, 1998). When behaviorally relevant features belong to different objects, can observers selectively attend to those features, or does object-based attention interfere with the selection process?

Duncan proposed that selective attention operates at the object-based level, limiting selection to one object at a time. Using a dual-task paradigm, subjects performed multiple perceptual judgments regarding features within one object or between two objects. Performance was better when the judgments regarded features belonging to the same object versus different objects (Duncan, 1984). Vecera and Farah (1994) later demonstrated similar divided attention costs—using Duncan's stimuli—regardless of whether the two objects were superimposed or separated in space, lending further support that object ownership limits perception. Ideally spatial separation could be better controlled for within the stimulus. In addition, the types of feature judgments made within and between objects were not the same leading to the concern that task demands may have been different across conditions.

Valdes-Sosa and colleagues improved the paradigm for studying object-based attention by superimposing two random dot fields which rotated in opposite directions (Valdes-Sosa, Cobo, & Pinilla, 2000). As is typical with a random dot kinemategram, each dot existed for a limited number of frames before being randomly redrawn, forcing surface segregation to rely on the global motion of the dots. Surface selection is therefore exclusively feature-based since dots belonging to either field were spatially intermingled. Subjects' reported the direction of two brief translational probes, which occurred in either surface and with a variable inter-probe interval. Performance in reporting the direction of the second probe dropped when it occurred in the other surface within 600 ms or less of the first. The authors described the first probe as an exogenous cue, which captured object-based attention decreasing the subjects' sensitivity to probes on the second surface. These results provide evidence that perception of objects is capacity limited. The authors describe the capacity limit as a "difficulty in switching attention rapidly between surfaces."

Multiple groups have argued that capacity is unlimited when dividing attention within an object but requires switching when attempting to divide attention between objects (Duncan, 1984; Blaser et al., 2000; Valdes-Sosa et al., 2000). In a dual-task paradigm where the observer is asked to divided their attention across multiple objects, an all-or-none switching model assumes that on a given trial the observer is constrained to select only one object at a time (Bonnel & Haftser, 1998; Sperling & Melchner, 1978) and therefore must guess when asked to recall the properties of a second object. Thus the all-or-none switching model predicts a negative trial-by-trial covariance (Bonnel & Prinzmetal, 1998) and a decrease in overall performance, known as a dual-task cost. In addition to the all-or-none switching model, there are limited-capacity parallel models that also predict a dual-task cost, but zero trial-by-trial covariance. One specific limited-capacity model is the fixed capacity model which maintains a fixed amount of information processing when attention is divided (Shaw, 1980). These models make specific predictions of dual-task performance from baseline single-task performance. In this paper we measured dual-task performance when attention was divided within versus between two surfaces and compared behavioral performance to the predictions of two common capacity models at either end of the capacity continuum: the unlimited-capacity parallel and the all-or-none switching models.

Most capacity models assume perfect selection, but a dual-task deficit could also arise from selective attention errors cause by distractor interference. The difficulty in dividing attention between objects may arise from interaction between feature-channels through *crosstalk* (Navon & Miller, 1987). If object-based attention facilitates the selection of all of an object's features, then task irrelevant features may interfere when attempting to select specific features from multiple objects (Davis, Driver, Pavani, & Shepherd, 2000). The interference due to crosstalk may increase when attention is divided between objects composed of competing features within a feature-dimension (e.g. different directions of motion). The prevalence of selection errors observed in our data lead us to purpose a new capacity model that takes into account crosstalk.

In order to measure the capacity of divided attention within and between objects, we measured accuracy when attention was divided across features within and between two superimposed transparent motion surfaces created from random dot kinemategram. Following in the tradition of Duncan and Valdes-Sosa, we chose to focus on accuracy rather than reaction times in order to test the predictions made by specific capacity models. Although our model does not make specific reaction time predictions, dual-task deficits may also manifest in slower reaction times when attention is divided between surfaces (Lamy & Egeth, 2002; Watson & Kramer, 1999). In addition, interference due to crosstalk has also been shown to affect reaction times (Navon & Miller, 1987; Treisman, Kahneman, & Burkell, 1983).

METHODS

Participants

Five subjects participated in this study, including the first author. All subjects gave written informed consent in accord with the human subject protocol at the University of Washington (Seattle, WA).

Apparatus

Stimuli were presented on a CRT monitor with a resolution of 1024 by 768 pixels, and viewed from a distance of 57 cm. Subject responses were collected by keyboard button presses. The code was written in MATLAB and presented using Psychtoolbox (Brainard, 1997; Pelli, 1997) on a computer running Windows

Stimuli

The stimuli consisted of two superimposed moving surfaces composed of randomly drawn dots with a unique color-motion conjunction, i.e. red-up and green-down or vice-versa, counter-balanced across sessions. Each surface was composed of 100 dots. To remove a potential depth cue, the depth order of overlapping dots (which dots occludes the other dot) was randomized. The diameter of the dots was 0.8 deg. The dots were confined to annulus with an inner diameter of 3 deg, and an outer diameter of 16 deg (see Figure 1B). A fixation plus was placed at the center of the annulus. The dots moved coherently at a rate of 8 degrees per second. Each dot was presented with a limited lifetime of 12 frames (200 ms), and was subsequently redrawn at a random position. The luminance of the green dots was reduced to match the luminance of the red dots at the maximum intensity of the red channel (33 cd/m^2). The (x, y) CIE 1931 *xyz* space coordinates for the red and green dot colors were (0.612, 0.331) and (0.279, 0.581), respectively. The monitor background was set to black with a luminance of less-than 1 cd/m^2.

Procedure

Prior to a block of trials, subjects received specific instructions regarding which feature, or pair of features, to attend in order to perform one or two detection tasks. There were a total of four conditions; two single-task and two dual-task conditions. For the *single-task motion* condition, subjects were cued to attend to the speed of one of the two surfaces (e.g. "attend to the speed of the upward moving red surface"). For the *single-task luminance* condition, subjects were cued to attend to the luminance of one of the two surfaces (e.g. "attend to the brightness of the red surface"). For the *dual-task, within-surface* condition, subjects were cued to attend to the speed and the luminance of one of the two surfaces (e.g. "attend to the speed AND brightness of the upward moving red surface"). For the *dual-task, between-surface* condition, subjects were cued to attend to the speed of one surface and the luminance of the second surface (e.g. "attend to the speed of the upward moving surface AND the brightness of the green surface").

The trial structure is schematized in Figure 1. Each trial began with a 1000 ms pre-trial-interval, consisting of a fixation plus centered on a black screen (Figure 1A). The stimulus movie followed for 1100 ms (Figure 1B). Stimulus "transients" occurred 500 ms after the onset of the moving surfaces and consisted of brief (100 ms) decrements in speed and/or the luminance of the dots within each surface (Figure 1B). On every trial, there was an independent 50% chance of a transient occurring in each of the four features. Thus, from zero to four stimulus transients occurred on each trial.

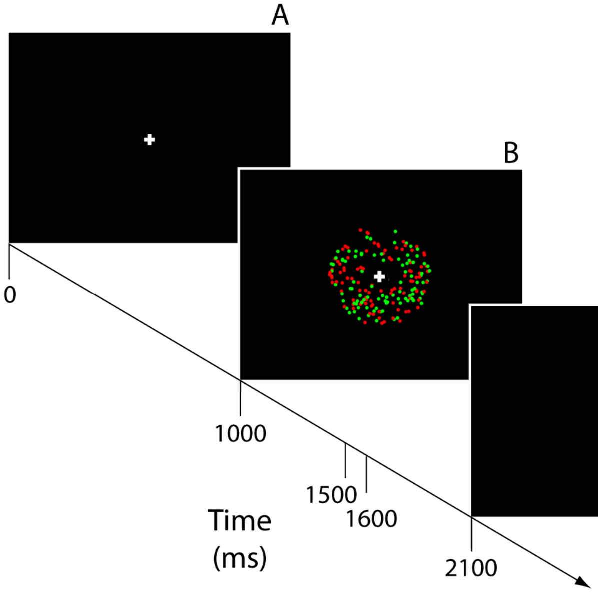

Figure 1 Trial structure. (A) Each trial commenced with a 1000 ms pre-trial-interval with a fixation-cross. (B) The stimulus then appeared for 1100 ms. At 500 ms, post stimulus onset, 0-4 100 ms luminance/motion transients could occur. (C) Following the stimulus, the observer made one or two yes/no responses without time pressure.

Following stimulus offset, subjects were queried to indicate whether or not a transient occurred in the cued feature(s) by pressing one of two buttons (a yes/no response). There was no time pressure. To help reduce response errors, the yes-no responses for the motion and luminance tasks were mapped to separate pairs of keyboard buttons; subjects used their left and right pointer fingers to perform the motion and luminance tasks (e.g. "did the speed of the upward moving surface decrease, yes (z) or no (x)?", and "did the brightness of the red surface decrease, yes (.) or no (/)". The order of report, whether motion or luminance was queried first, was pseudo-randomized across trials to prevent response biases.

Subjects practiced all four experimental conditions over the course of two to three one-hour sessions until they reported feeling comfortable with the task. Experimental data was then collected over two one-hour sessions. The magnitudes of the speed and luminance decrements were adjusted during the training sessions to ensure single-task performance levels above 80% correct but bellow ceiling. Intensities decrements were chosen separately for each surface (Table 1). For the motion task, subject 3 exhibited much higher sensitivity for upward motion decrements then downward motion decrements. For the luminance task, 4 subjects (2–5) exhibited slightly higher sensitivity for the luminance decrements in the green field then in the red field.

During the experimental sessions the magnitude of the speed and luminance decrements were held constant. Subjects performed blocks of 32 trials, preceded by specific attention instructions. Each attention condition was grouped into sets of four blocks. For example, for the single-task motion subjects alternated between attending upward motion for 32 trials, and downward motion for 32 trials. After four blocks (128 trials total), the subject began a new (randomly selected) cue condition. In this manner, 256 trials were collected for each of the four cue conditions in each one-hour sessions for a total of 512 trials per cue condition.

	Speed decrement (%)		Luminance decrement (%)	
Subject	Upward	Downward	Red field	Green field
1	43	43	36	36
2	36	36	26	29
3	63	50	34	38
4	45	45	33	36
5	26	26	27	36

Table 1 Percent of speed or luminance decrement from baseline intensity (8 deg/sec and 33 cd/m^2) for transient events. The percentages are tabulated for each surface feature within each subject.

Analysis

We collapsed performance across surfaces, because we were not interested in performance differences within a feature-dimension (e.g. upward vs. downward motion or red vs. green). Behavioral performance was analyzed at several different levels. For a coarse analysis of performance, we averaged across the hit rate and correct rejection rate to compute a percent correct for each condition. Going further, we analyzed the joint dual-task performance for signs of independence between tasks. At the finest level of analysis, we compared hit rates to false alarm rates conditionalized on each type of stimulus transient. Finally, we fit a parametric model to the most informative of the three conditionalized response distributions (*see* Modeling *below*)

Error bars

When plotting averages across subjects, error bars error bars encompass ±1 standard error of the mean (n=5). When plotting individual subject data, we resampled our data (with replacement) 10,000 times, calculating the sample mean after each iteration (Wichmann & Hill, 2001). Error bars enclose ±1 standard deviation of the sampling distribution (±34.14%).

RESULTS

We began by comparing dual-task performance to single-task performance in an attention operator characteristic (AOC) plot (Figure 2) (Sperling & Melchner, 1978). AOC plots are generated by plotting dual-task performance levels against one-another. To ease comparison between the single-task and dual-task conditions, single-task performance is plotted along the axes. The effect of divided attention within versus between-surfaces is readily apparent in the AOC which shows mean performance across the five subjects.

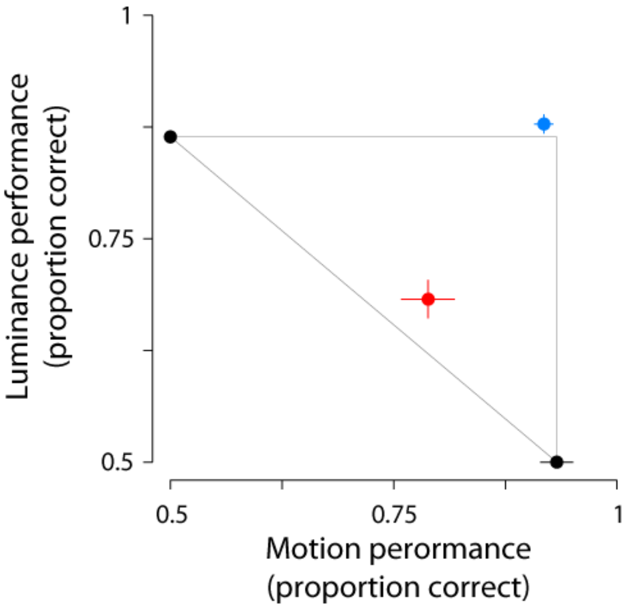

Figure 2 Performance on the motion task (abscissa) is plotted against performance on the luminance task (ordinate). Between-subject average single-task performance is plotted in black on the axes; dual-task performance, when attention was divided within a surface, is plotted in blue; dual-task performance, when attention was divided between-surfaces, is plotted in red. Error bars encompass ±1 standard error of the mean between-subjects (n=5). Grey lines extend from the single-task performance levels to aid in their comparison to dual-task performance levels.

Detecting a change in both the motion and luminance within a single surface yielded little to no deficit compared to detecting a change in either feature alone. The average within-subject difference between single-task performance and within-surface dual-task performance (0.02±0.01 for the motion task, and −0.02±0.02 for the luminance task) was not statistically significant (t(4)=1.46 & 1.45; p>0.05). In contrast, detecting a change in the motion of one surface and the luminance of the other surface resulted in a significant deficit compared to single-task performance levels (−0.14±0.02 & −0.18±0.017). The difference was statistically significant for both tasks (t(4)=7.05 & 10.26; p<0.01). The pattern of dual-task performance was consistent across all five subjects (Figure 3).

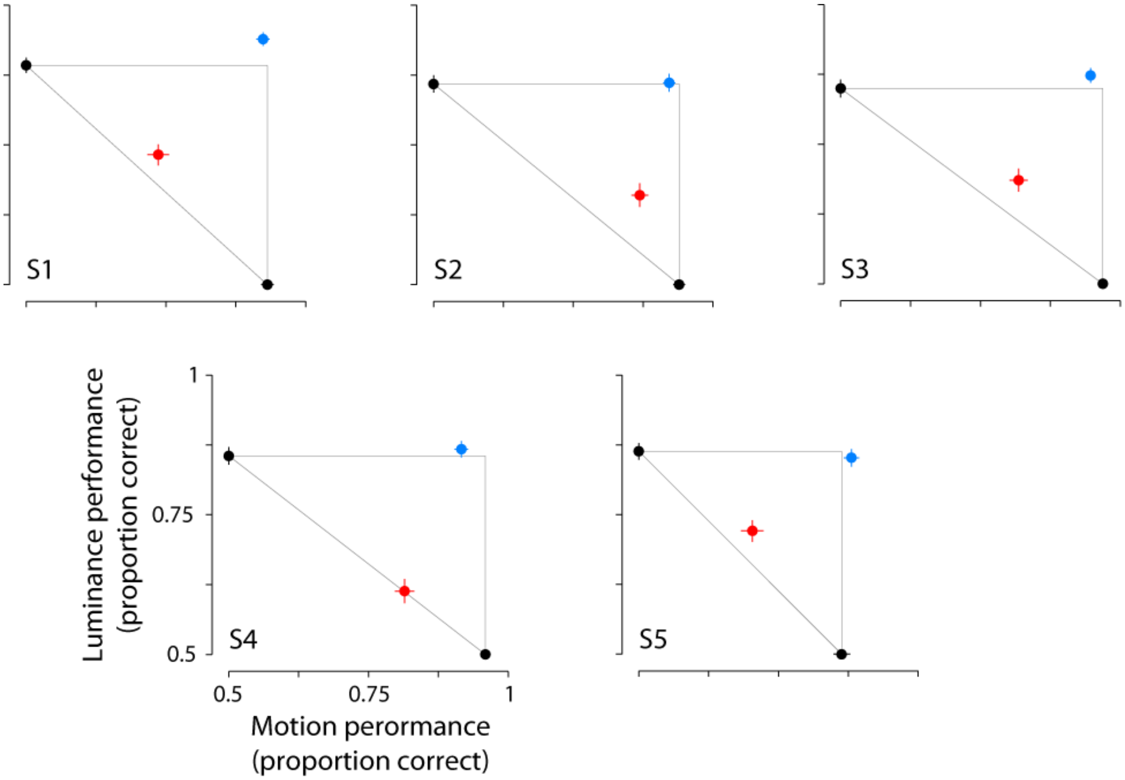

Figure 3 Individual subject performances are summarize by separate AOC plots. Error bars enclose ±1 standard deviation of the bootstrapped sampling distribution (see *Methods: Error bars*).

To verify that there were no memory or order effects resulting from the order in which the two tasks were performed in the dual-task conditions, we separated performance on the basis of response order (Figure 4). The average within-subject differences between performance in the motion task when motion was queried first versus when color was queried first (−0.013±0.005 for the within-surface condition, and −0.009±0.011 for the between-surface condition) was not statistically significant (t(4)=2.36 & 0.90; p>0.05). Likewise, the differences between performance in the color task when color was queried first versus when motion was queried first (0.00±.02 & 0.01±0.02) was not statistically significant (t(4)=0.12 & 0.71; p>0.05).

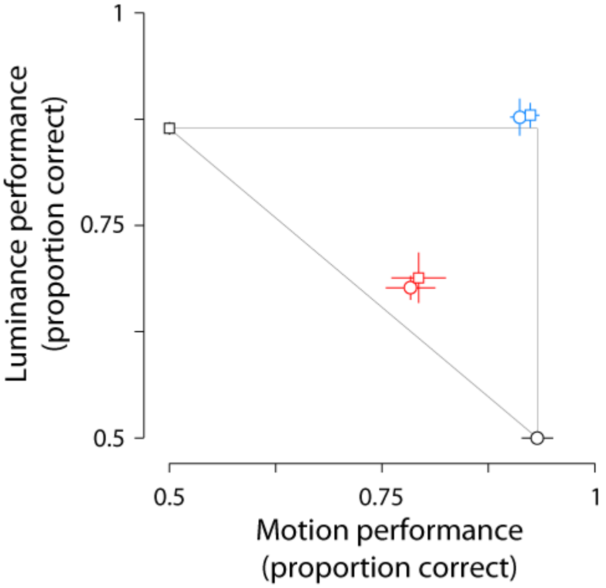

Figure 4 AOC plots, averaged across subjects, after separating dual-task performance on the basis of response order. Squares denote performance over trials in which the luminance task was queried first, and circles denote performance over trials in which the motion task was queried first.

MODELING

We next explored the predictions of two simple non-parametric models at either end of the processing-capacity continuum. The unlimited-capacity sharing model assumes that each task was performed independent of the other task, whereas the all-or-none switching model assumes that only one task can be performed at a given time.

Unlimited Capacity Model

There exists a range of models over which the amount of interference between two concurrent tasks varies. At one end of the spectrum resides the unlimited-capacity parallel model, which assumes independent and non-interfering task performance. Given unlimited capacity, the joint probability of two correct responses in the dual-task condition is simply the product of the two single task performance levels:

$$p(m \,\&\, l)_{dual} = p(m)_{\text{single}} * p(l)_{\text{single}} \quad (1)$$

where m and l represent the motion and luminance tasks, and $p(x)$ represents the probability of a correct response on task x. The unlimited-capacity sharing model predicts that dual-task performance should fall at the independence point at the intersection of the single-task performance levels.

The results from the within-surface dual task falls very close to this intersection for all five subjects (Figure 3), which suggests that there is not a capacity limit for dividing attention across features within a single object.

All-or-none Switching Model

On the other end of the spectrum resides the all-or-none switching model, which assumes that only one task can be carried out at a given time. Consequentially, the model predicts a negative trial-by-trial correlation in dual-task performance because the observer can only be in one attentional state at a time. When attention is directed to one task the observer must guess on the other. This model results in two contingency tables, one for each attention state. Since subjects were given no priority instructions, we assumed an even mixture of the two attention states across trials. If the observer attended to motion on half of the trials while guessing on the luminance task, and attended to luminance on the other half of the trials while guessing on the motion task, the joint probability of getting the luminance and motion task correct is:

$$p(m\&l)_{dual} = \frac{0.5 \times p(m)_{single} + 0.5 \times p(l)_{single}}{2} \tag{2}$$

The all-or-none switching model predicts a trade-off between the two tasks confining dual-task performance along the negative diagonal connecting the two single-task performance levels (see Figures 1 and 2). Or more specifically, halfway between chance and single-task performance, given an equal combination of unattended trials in which the subject guessed attended trials in which the subject performed at the single-task level. For each of the five subjects (Figure 3), dual-task performance is close to the negative diagonal when attention was divided between-surfaces. This seems at first like strong support for the all-or-none switching model for predicting dual-task deficits for dividing attention across surfaces. However, a further analysis shows that this model cannot describe the results. A key feature of the all-or-none switching model is that there should be a negative covariance between trial-by-trial performances, since attention to one task forces the subject to guess on the other. Note that while the AOC plots in Figures 1 and 2 provide a useful graphical summary of performance in divided attention experiments, because performance is collapsed across trials, trial-by-trial covariance cannot be observed in these plots.

The amount of negative covariance predicted by the all-or-none switching model depends on the single-task performance level. Plotted in Figure 5 is the observed dual-task covariance between the motion and luminance task, along with the prediction curve of the all-or-none switching model as a function of single-task performance (Equation 3):

$$(3) \qquad cov\left(PC_{single}\right) = p(m\&l) - [p(m\&l) + p(m\&{\sim}l)] \times [p(m\&l) + p({\sim}m\&l)]$$

where *p(m&l)* corresponds to the joint probability of correct response on both task described by Equation (2). To reduce the dimensionality of the space we set $p(m)_{single}$ equal to $p(l)_{single}$ for each point along the abscissa. The *"~"* in Equation (3) signifies the probability of an incorrect response—otherwise one minus the probability of a correct response. The between-subject average covariance (0.001 ± 0.003) was statistically indistinguishable from zero (t(4)=0.56; p>0.05). In addition, there was plenty of power to reject the prediction of the all-or-none, switching model. The average within-subject difference between the observer covariance and the model prediction (0.041±0.002) was statistically significant (t(4)=16.93; p<0.001).

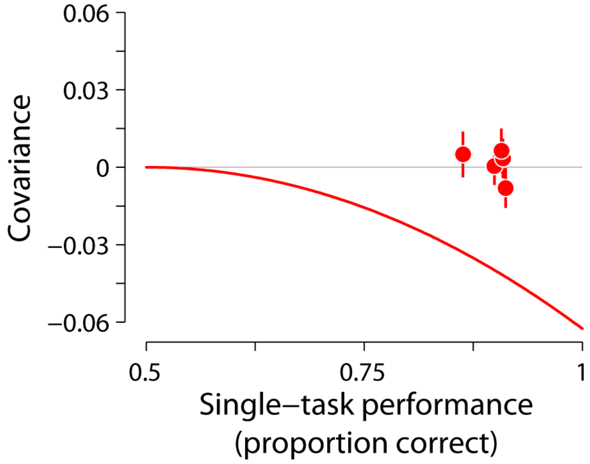

Figure 5 Single-task performance is plotted on the abscissa, and trial-by-trial covariance between the motion and luminance task is plotted on the ordinate. The red data points correspond to observed single-task performance and dual-task covariance for when attention was divided between surfaces. To reduce the dimensionality, single-task performance was averaged across the two tasks. Error bars enclose ±1 standard deviation of the bootstrapped distribution (horizontal error bars fall within the extent of the data points). The red curve traces the covariance predicted by the all-or-none switching model as a function of single-task performance.

Conditionalizing Responses on Target and Distracter Transients:

Dividing attention across surfaces led to a decrease in performance, but without a corresponding negative trial-by-trial covariance. Instead, some other mechanism besides switching must be causing this performance deficit. One possibility is a phenomenon called *crosstalk* (Navon & Miller, 1987) which is when subjects inadvertently respond to the presence of distractors. For example, subjects may be more likely to respond 'yes' to an upward speed decrement when a speed decrement occurs on the uncued downward moving surface.

To further investigate the crosstalk hypothesis, we conditionalized hits and false alarms on the presence or absence of uncued transients, or *distractors*. If there is crosstalk between channels, then distractors will not be properly filtered and an increase in hits and false alarms on trials containing distractors would occur. For each task there were three categories of distractors: same feature / different surface, different feature / same surface, and different feature / different surface. For example, when the motion of surface 1 was cued a distractor transient may occur in the, motion of surface 2, the color of surface 1, and/or the color of surface 2.

To begin, we considered distractors within the same feature-dimension (Figure 6A,B), e.g. for the motion task, a downward motion transient on trials where upward motion is cued (Figure 6A). The presence of these distractors increased the proportion of 'yes' responses, and therefore both the proportion of hits and false alarms (solid points)—for both tasks and across all conditions—relative to trials containing no distractors (open points). For the motion task (Figure 6A), the average within-subject increase in proportion of yes-responses (0.06±0.02 for the single-task, 0.07±0.01 for the dual-task within, and 0.17±0.03 for the dual-task between), was statistically significant across all conditions ($t(4)=3.40$, 5.06, & 5.93; $p<0.05$, <0.01, & <0.01). For the luminance task (Figure 6B), the average within-subject increase in proportion of yes-responses (0.23±0.02, 0.20±0.02, & 0.31±0.03), was statistically significant across all conditions ($t(4)=13.74$, 8.42, & 10.13; $p<0.01$).

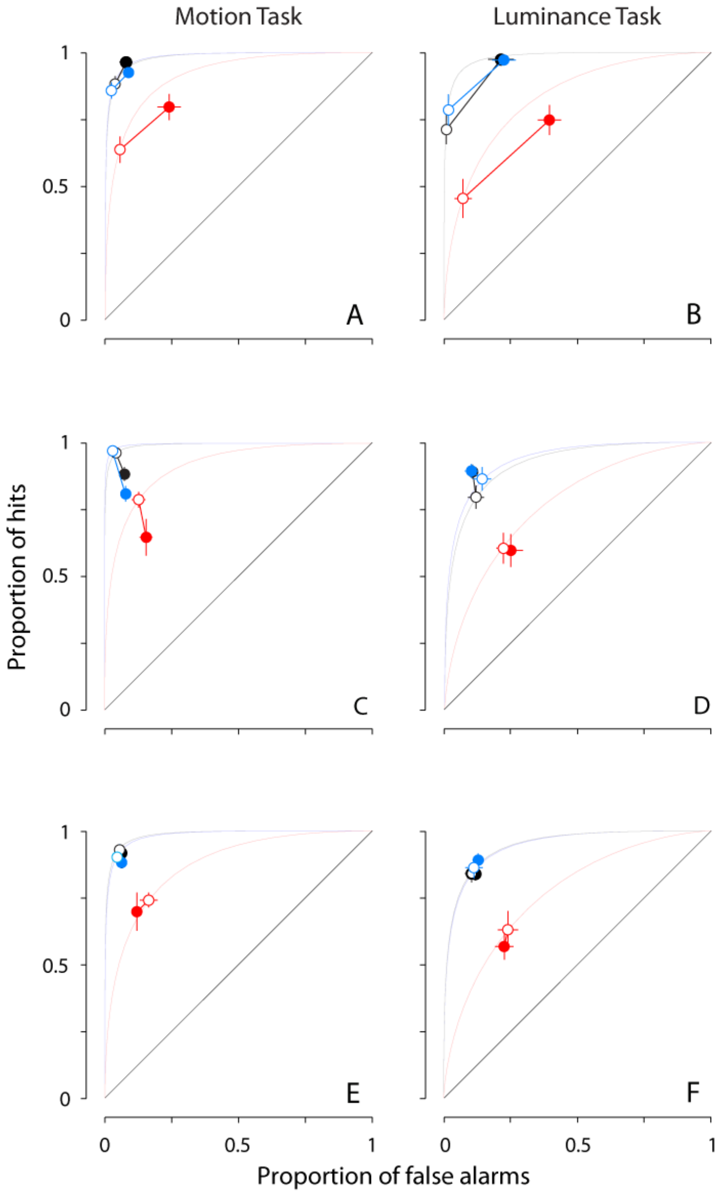

Figure 6 Between-subject average false alarm rate (abscissa) and hit rate (ordinate) contingent on distractor transients. Open points correspond to trials with no distractor transients, and filled points correspond to trials with distractor transients. Isosensitivity curves (assuming zero bias) are drawn through the open points (Green & Sweets, 1974). False alarm and hit rates for (A) the motion task, contingent on motion transients in the other surface; (B) the luminance task, contingent on luminance transients in the other surface; (C) the motion task contingent on luminance transients in the same surface; (D) the luminance task contingent on motion transients in the same surface; (E) the motion task contingent on luminance transients in the other surface; (F) the luminance task contingent on motion transients in the other surface.

Based on signal detection theory, (Green & Sweets, 1974), we drew isosensitivity curves (constant d-prime) through the open points to help visualize changes in sensitivity from changes in response bias due to the distractors. For the motion task, the difference in d-prime (−0.4±0.4, 0.3±0.3, & 0.4±0.3) was not statistically significant (t(4)=0.93, 1.22, 1.57; p>0.05). For the luminance task, the difference in d-prime (−0.2±0.3, 0.4±0.3, & 0.5±0.2) was also not statistically significant (t(4)=0.61, 1.54, 2.10; p>0.05).

Next, we considered distractors within the other feature-dimension on the same surface (Figure 6C,D), e.g. for the motion task, a luminance transient in the same surface as the cued motion direction. Luminance transients decreased the accuracy of motion judgments within the same surface across all three cue conditions (Figure 6C). The decrease in sensitivity (by a d-prime of 1.1±0.2, 1.8±0.2, & 0.6±0.1) was statistically significant across all three conditions (t(4)=4.68, 10.00, & 4.17; p<0.01, 0.01, & 0.05). This suggests that luminance transients masked target-motion transients within the same surface. The complementary effect was not observed in the luminance task (Figure 6D); motion transients did not appear to mask target luminance transient within the same surface. The change in sensitivity to luminance targets when motion distractors occurred in the same surface (−0.5±0.2, −0.3±0.2, & 0.1±0.1) was not statistically significant (t(4)=2.68, 1.54, & 0.83; p>0.05).

Finally, we considered the effect of distractors within the other feature-dimension, on the other surface (Figure 6E,F), e.g. for the motion task, a luminance transient in the surface moving in the uncued motion direction (Figure 6E). These distractors had no measurable effect on responses in either task across all three cue conditions. The change in sensitivity for motion transients when luminance distractors occurred (Figure 6E) in the other surface (0.1±0.1, 0.3±0.2, & −0.1±0.2) were not statistically significant (t(4)=1.21, 0.15, 0.74; p>0.05). Likewise, the change in sensitivity for luminance transients when motion distractors (Figure 6F) occurred in the other surface (0.1±0.1, 0.00±0.1, & 0.1±0.1) were not statistically significant (t(4)=1.12, 0.03, 1.46; p>0.05).

The two main effects captured by the AOC plot (Figure 5) are also present in all of the ROC plots (Figure 6). First, the conditionalized responses are similar between the single-task (black) and dual-task within (blue) surface conditions. Second, when attention was divided between surfaces (red), there was a decrease in sensitivity compared to the other two conditions (i.e. the red points fall closer to the diagonal than the other two conditions). Distractors within the same feature-dimension increased the proportion of yes-responses across both task and in all three conditions (Figure 6A,B). Decrements in luminance masked decrements in motion (decreasing sensitivity) within the same surface (Figure 6C), but not vice-a-versa (Figure 6D). However, luminance events in the other surface had no masking effect on the motion task (Figure 6E).

A Limited-capacity Sharing Model with Crosstalk

One way to describe how crosstalk interferes with selection is to imagine that some proportion of the output from the distractor channel is leaked into the output of the target channel. Poor selection can be exemplified by the extent to which the probability of a yes-response is greater when: 1) a distractor alone occurred compared to no transients at all, and/or 2) both a target and distractor occurred compared to a target alone.

We formalized this crosstalk concept into a model called *the limited-capacity sharing model with crosstalk*. The term, *limited capacity* refers to the fact that we allowed sensitivity to vary freely across conditions, in contrast to the specific limited-capacity model which assumes a fixed rate of information processing (Shaw, 1980). The term *sharing* refers to the assumption that both tasks are performed independently but with limited capacity. The model begins with an encoding stage: each feature is encoded by an independent sensory channel, the output of which is a normally distributed random variable. We assumed that on a transient-absent trial the output of the channel was drawn from a “noise” distribution: a normal distribution with a mean of zero and a standard deviation of one. On transient-present trial the output of the channel was drawn from a “signal” distribution: a normal distribution with a mean greater than or equal to zero and a standard deviation of one. Finally, we assumed that the sensitivity of the two motion/luminance channels were the same (e.g. the sensitivity to upward and downward motion is equivalent).

To illustrate the model, consider the motion-task for a given trial in which upward motion is cued (Figure 7). Figure 7A depicts the probability density functions (PDF) for the outputs of the upward and downward motion channels (above and below respectively). The random variable x_1 denotes the internal evidence for an upward motion transient (target), and the random variable x_2 denotes the internal evidence for a downward motion transient (distractor). In the absence of a transient, the output of either channel is drawn from the noise distribution (with a mean equal to zero). On trials containing a target or distractor transient the output is drawn from the signal distribution (with a mean shifted from zero). Based on the assumption that attention has no effect on the stimulus encoding stage, the mean of the signal distribution for target and distractor transients are the same. The mean of the signal distribution is equal to the sensitivity (d-prime). As sensitivity decreases the overlap between the two distributions increases making the perceptual discrimination between stimuli present/absent more difficult.

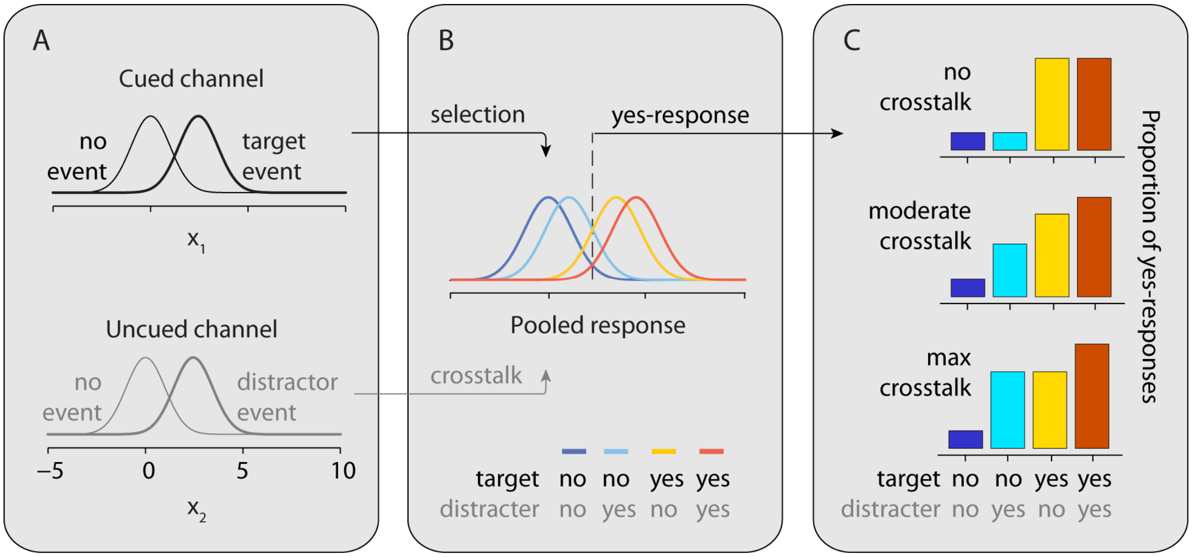

Figure 7 Schematic for the limited-capacity sharing model with crosstalk. (A) The output of two opposing channels (e.g. upward and downward motion selective channels) is assumed to be normally distributed random variables with unit standard deviation. Trials containing a target transient (above) generate a larger mean response within the cued channel, represented by the shifted PDF. Trials containing a distractor transient (below) generate a larger mean response within the uncued channel, represented by the shifted PDF. A yes-no decision is based on the output of the cued channel plus some leak, or crosstalk, from the uncued channel. There are four possible target/distractor combinations resulting from the two channels. Based on the amount of crosstalk, the means and standard deviations of the four evidence distributions will vary. (B) The four pooled response distribution are shown for a moderate level of crosstalk. The distributions are colored coded as follows: no target or distractor transient (blue), distractor transient alone (cyan), target transient alone (yellow), and both target and distractor transients (red). The dotted gray line represents a possible decision boundary, or criterion, above which a yes-response is made. (C) The amount of crosstalk will change the conditional probability of a yes-response. Three example distributions are shown for a fixed sensitivity and response criterion given, no crosstalk (top), moderate crosstalk (middle), and max crosstalk (bottom).

Following stimulus encoding, a decision is based on the output of the cued channel (x_1), plus some amount of leak, or crosstalk, from the uncued channel (x_2). The amount of crosstalk is controlled by a gain term. If selection were perfect then the value of the gain parameter would equal zero and the distractor would have no effect on the decision. If the subject was unable to differentiate the target from the distractor—a complete failure of selective attention—then the gain parameter would equal one. A yes-response is made if the pooled output of the two channels is greater than a criterion value (Figure 7B). The proportion of yes-responses increases as a function of crosstalk (Figure 7C). If selection were perfect (a crosstalk gain parameter equal to zero), then distractors should have no effect on responses (Figure 7C, top). Given a moderate level of crosstalk (gain=0.5), distractors will increase the proportion of yes-responses (Figure 7C, middle). Given a complete failure of selective attention (gain=1.0), the proportion of yes-responses given a distractor alone will equal the proportion of yes-responses given a target alone, and will increase to the combined probability of a yes-response when both a target and distractor occur (Figure 7C, bottom).

The model contains three parameters; sensitivity, which defines the mean channel output corresponding to a transient-present trial (signal distribution); a gain term, which controls the amount of leak from the uncued channel; and a decision criterion, which determines how large the pooled output from the two channels must be in order to produce a yes-response. We used a maximum likelihood procedure to estimate the parameter values that yielded the greatest probability of generating our observed data set. In order to take full advantage of the information in our data set, we divided trials into four categories based on the pairwise combination of target and distractor transients, and tallied the number of yes-responses in each category. We then fit the model to these four yes-response probabilities. The motion and luminance tasks for each of the three cue conditions were fit separately.

To visualize the model predictions to the data, we re-plotted in Figure 8 the values from each pair of ROC points from Figure 6A,B on a common axis – the proportion of yes-responses. Three general patterns are immediately apparent when inspecting the proportion of yes-responses for each task across the three cue conditions. (1) The distribution of yes-responses was nearly identical between the single-task condition and the dual-task, within-surface condition. (2) Crosstalk was more evident for the luminance task than for the motion task across all three cue conditions (compare the proportion of yes-responses with and without distractors: cyan vs. blue & red vs. yellow). (3) Performance dropped, and selection errors became more prevalent when attention was divided between-surfaces.

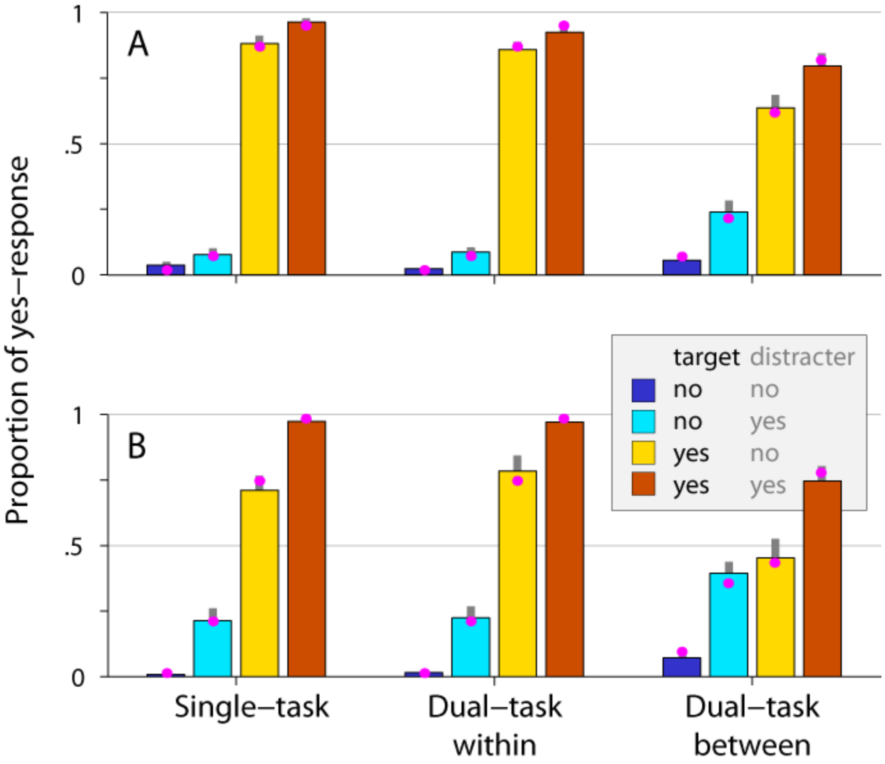

Figure 8 Bars represent the between-subject average proportion of yes-responses conditionalized on the four target/distractor combinations (re-plot of data from Figure 6A,B). Error bars enclose ±1 standard error of the mean. Model predictions are plotted in cyan. Because single-task and dual-task within conditions were simultaneously fit, the cyan points are identical between the two cue conditions. (A) Proportion of yes-responses in the motion task given—from blue-to-orange—no motion transients in either surface; one motion transient in the uncued surface (distractor); one motion transient in the cued surface (target); motion transients in both surfaces (target + distractor). (B) Proportion of yes-responses in the luminance task given—from blue-to-orange—no luminance transients in either surface; one luminance transient in the uncued surface (distractor); one luminance transient in the cued surface (target); luminance transients in both surfaces (target + distractor).

The difference in the observed probability distributions between the single-task and the dual-task, within-surface conditions (Figure 8) was statistically indistinguishable (chi-squared(3,4)=51.81 & 28.87 for the luminance and motion task respectively; p>0.05). Fitting the model separately to these two conditions improved the fit by less than 3% (increase in maximum likelihood) for the motion condition and less than 1% for the color condition. Thus, we reduced our parameters by fitting the combined data for the two conditions (single-task and dual-task within), hitherto referred to as the baseline condition. None of the residual differences between the model predictions and the observed proportion of yes-responses were not statistically different from zero (p>0.05 for all 24 t-tests) (Figure 9).

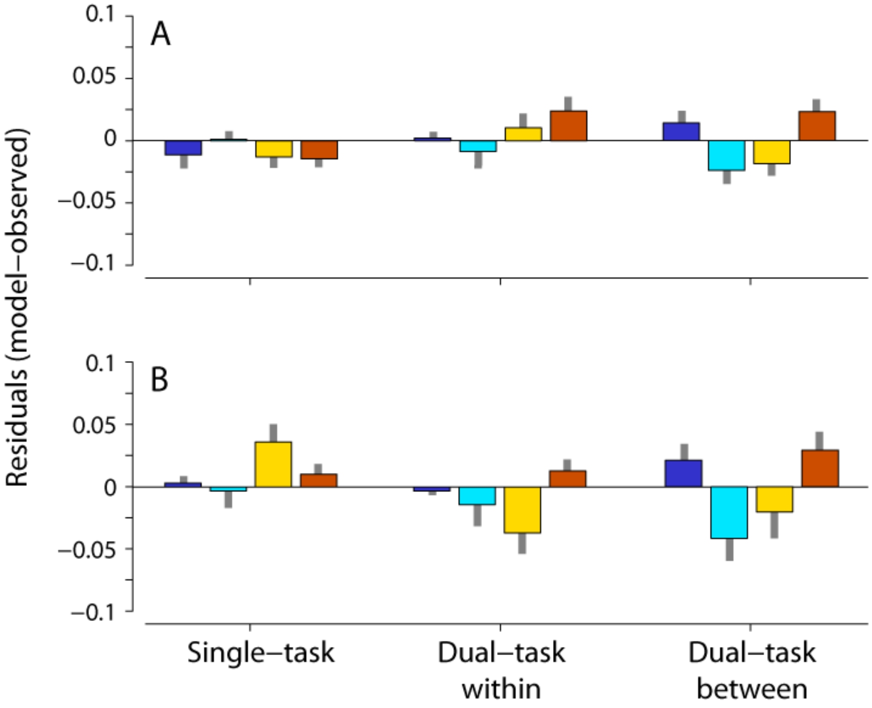

Figure 9 Between-subject residual error; difference scores between the limited-capacity sharing model predictions and the observed proportion of yes-responses for the motion task (A) and luminance task (B).

The average parameter values for the baseline and the dual-task, between-surfaces conditions, are displayed in Table 2. Considering the baseline condition alone (Table 2, first column), the model describes behavioral performance as follows. First, detection sensitivity was high for both tasks (d-prime of 3.4±0.2 for the motion task and 3.8±0.2 for luminance task). Second, the crosstalk gain parameter determines how well the subjects were able to select the cued feature and ignore distractor transients within the same feature-dimension. The crosstalk gain parameter was significantly greater than zero in both cases ($t(4)=6.95$ & 17.75, $p<0.01$), suggesting that even in the baseline condition subjects were not able to completely filter out distractors. There was more crosstalk in the luminance task then in the motion task (0.50±0.03 versus 0.17±0.02, for the luminance and motion task respectively). This difference is reflected in the data by the increase in the false alarm rate when a distractor transient occurred (0.21±0.05 vs. 0.04±0.02, Figure 8 difference between cyan and dark blue bars), and an increase in the hit rate when a distractor transient co-occurred with a target transient (0.26±0.06 vs. 0.08±0.02, Figure 8 difference between orange and yellow bars). Finally, the response criterion determines the trade-off between false alarms and misses. A response criterion equal to half an observer's sensitivity—zero bias—predicts an equivalent false alarm and miss rate. The higher the criterion—a conservative bias greater than zero—the more sensory evidence the observer requires to make a yes-response. A conservative observer with a high criterion will commit more misses in order to avoid false alarms. Subjects tended to be conservative in both tasks (bias of 0.45±0.04 & 1.02±0.20, for the motion and luminance tasks respectively), committing fewer false alarms than misses. This parameter is reflected in the data by a very low false alarm rate when no distractor occurred (0.037±0.012 & 0.009±0.004, Figure 8 dark blue bars).

	Baseline		Dual-task between		Log parameter ratio	
	motion	*luminance*	*motion*	*luminance*	*motion*	*luminance*
Sensitivity	3.4 (0.2)	3.8 (0.2)	2.2 (0.3)	1.6 (0.1)	−0.21* (0.04)	−0.37* (0.03)
Crosstalk gain	0.17 (0.02)	0.50 (0.03)	0.39 (0.07)	0.84 (0.09)	0.35* (0.09)	0.21* (0.04)
Response criterion	2.14 (0.08)	2.92 (0.22)	1.8 (0.1)	1.9 (0.2)	0.20* (0.03)	0.03 (0.10)

Table 2 Average between-subject parameter values for baseline condition (combined single-task & dual-task within), dual-task between surface condition, and the $\log_{10}$ of the ratio between the two (dual-task between divided by baseline). Standard error of the mean are printed below in parentheses. Parameter values were fit to each subjects' data set using a maximum likelihood procedure that maximized the likelihood of the observed conditional proportion of yes-responses. The sensitivity and response criterion are in units of d-prime. The crosstalk gain parameter ranges between zero (no crosstalk) and one (maximal crosstalk). Within-subject t-tests were conducted on the log of the parameter ratios; red font denotes a log ratio significantly greater than zero ($p<0.05$).

The effect of dividing attention between surfaces is captured in the model by the ratio of the parameter values for the dual-task, between surface and baseline conditions. The $\log_{10}$ of this ratio is tabulated in the third column of Table 2. Dividing attention between surfaces results in: (1) a significant decrease in sensitivity shown by a log sensitivity ratio of -0.21 ± 0.04 for the motion task, and -0.37 ± 0.03 for the luminance task ($t(4)=4.96$ & 12.97, $p<0.01$), plotted on the left in Figure 10; (2) A significant increase in crosstalk shown by a log of the crosstalk gain ratio of 0.35 ± 0.09 for the motion task and 0.21 ± 0.04 for the luminance task ($t(4)=4.09$ & 5.30, $p<0.05$), plotted on the right in Figure 10. In addition, there was also a conservative shift in bias (corresponding to a change in d-prime units of 0.26 ± 0.04 & 0.05 ± 0.21) shown by a log criterion ratio of 0.20 ± 0.03 for the motion task, and 0.03 ± 0.10 for the luminance task. However, this effect was only significant for the motion task ($t(4)=5.8$, $p<0.05$), and not for the luminance task ($t(4)=0.29$; $p>0.05$). For the luminance task, 2 of the 5 subjects showed a liberal shift in bias.

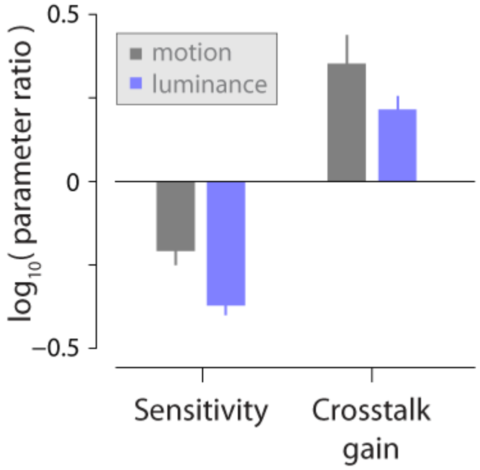

Figure 10 Log$_{10}$ of the ratio between the dual-task between and the combined single/dual-task within parameters (see Table 2) for the motion task (gray) and luminance task (light blue).

An observer may choose to implement a transient detection strategy by ignoring the cue and responding to transients in either surface (e.g. upward or downward speed changes for the motion task; or luminance changes across the red or green dots for the luminance task). When asked to divide attention between-surfaces, did the subjects choose to pursue a transient detection strategy or were they unable to simultaneously select features from competing surfaces? A transient detection strategy would result in poor performance since observer's false alarm rate would equal their hit rate. Such a strategy would place an observer's performance on the negative diagonal in the AOC plots (Figure 3), equivalent to ignoring one of the two cues (as predicted by the all-or-none switching model). Only one subject (S4) demonstrated that level of dual-task deficit. The other four subjects outperformed the theoretical low limit of dual-task performance. In addition, a pure transient strategy would result in a crosstalk gain parameter value of 1. The maximum likelihood estimate for the crosstalk gain parameter was below 1 for both tasks (0.84±0.09 & 0.39±0.07). This suggests that even though selection was poor, subjects were, at the very least, attempting to ignore the distractor transients.

To summarize, the proportion of yes-responses conditionalized on target and distractor transients (within the same feature-dimension) were statistically identical between the single-task and the dual-task, within surface conditions. Our limited-capacity model with crosstalk adequately fit our data set (Figure 8) with no consistent residual error (Figure 9). Dividing attention between surfaces resulted in a dual-task deficit described by the model as a decrease in detection sensitivity paired with an increase in crosstalk (Figure 10). The increase in crosstalk in the luminance task approached a complete failure of selective attention.

Overall, performance was statistically indistinguishable between the single-task conditions (motion or luminance task) and the dual-task within surface condition (motion and luminance task). In contrast, dividing attention between surfaces to perform the motion and luminance tasks resulted in a significant dual-task deficit. Performance across the two tasks was statistically independent (uncorrelated), contrary to the predictions of an all-or-none switching model. In addition, distractors within the same feature-dimension on the other surface increased the proportion of yes-responses, indicative of crosstalk. Although crosstalk was observed in all conditions, it was greatest when attention was divided between surfaces (Figure 6). Although distractors in the other feature on the other surface were successfully filtered (Figure 8), luminance distractors masked motion transients within the same surface (Figure 7). We constructed a limited-capacity sharing model that includes a crosstalk gain parameter to account for crosstalk within the feature-dimension. Our model successfully fit the observed proportion of yes-responses conditionalized on targets and distractors.

DISCUSSION

We used transparent motion to investigate the capacity limits in divided attention within and between objects. Transparent motion provides a useful stimulus for studying object-based attention because it allows for multiple surfaces to be superimposed, isolating object- and feature-based selection from spatial selection. Despite randomizing the depth order of overlapping dots so that neither surface consistently occluded the other, it is possible that the visual system segments the two surfaces in perceived depth. It is rare for natural stimuli to slide past one another without interacting and thus the visual system may interpret superimposed dot fields as occupying separate planes in depth. Despite this possibility, random dot motion remains a good way of controlling for spatial attention. In addition our paradigm possessed two key features that are important for studying attention. First, our stimulus was identical across conditions. Therefore, changes in performance result from capacity-limits in dividing attention rather than sensory encoding effects. Second, the task was held constant between the two divided attention conditions—in both cases the observer performed a motion and luminance task—so changes in performance result from capacity limits in object-based attention rather than task-based effects.

Understanding the effects of divided attention on the processing of multiple features within and between objects is central to models of object-based attention. In this study, we have shown evidence that all features within a relevant surface can be selected with unlimited capacity for the detection of motion and luminance transients. Unlimited capacity for features within an object has also been shown for a wide range of stimuli including: the tilt & texture of a line (Duncan, 1984), the color and shape of a letter (Bonnel & Prinzmetal, 1998), and the orientation, spatial frequency, and color of a Gabor patch (Blaser et al., 2000). This evidence further supports the hypothesis that object-based attention allows unlimited capacity processing of multiple features within an object.

We also showed evidence that dividing attention across two surfaces reduced performance. A deficit in dual-task performance when dividing attention across objects has been reported in a variety of studies across a range of superimposed stimuli (Duncan, 1984; Blaser et al., 2000; Scholl, 2001). Duncan proposed that object-based selection is all-or-none, limiting selection to one object at a time. If selection were all-or-none, then we should expect a negative correlation between a subjects' motion and luminance performance. Instead, we observed no significant negative correlation, and there was enough power (based on high single-task performance levels) to reject the prediction of the all-or-none switching model (Figure 5). Duncan's stimuli involved some degree of spatial segregation which could have contributed to the all-or-none switching effect that he argued for. Consistent with this explanation, subjects' dual-task performance was negatively correlated when they were cued to attend to the shape of one object and the color of a second displaced object (Bonnel & Prinzmetal, 1998). Because we controlled for spatial separation we cannot directly compare our results to Bonnel & Prinzmetal's.

Valdes-Sosa also argued that object-based attention was all-or-none (Valdes-Sosa et al., 2000) using a transparent motion stimulus similar to the one used in this study. However, they manipulated the temporal asynchrony (SOA) between the presentations of the two target probes. It is possible that the first probe exogenously captured attention, leading to the prioritization of information processing from that channel (Shomstein & Yantis, 2004). The time given by the SOA may have encourages subjects to switch attention between surfaces. We presented targets and distractors simultaneously in order to discourage such a strategy. Our limited-capacity sharing model could be modified to account for differences in task priority, but since we did not manipulate task priority in this study, our model remains ambivalent to the possibility that attention can be flexibly allocated between tasks (but see, Bonnel & Prinzmetal, 1998; Sperling & Melchner, 1978, for evidence of flexible allocation).

For each task there were three types of possible distractors. In order to address the influence of these distractors on responses, we analyzed hits and false alarms, contingent on each type of distractor. Begin by considering transient in different features. Distractors within the same surface had an asymmetrical effect on responses. Luminance transients interfered in the detection of motion transients (Figure 6C), but motion transient had no effect on sensitivity to luminance transients (Figure 6D). The masking of motion transients by luminance was observed across all three cue-conditions suggesting that a one way sensory interaction occurs between luminance and motion when a dot field simultaneously decreases in speed and luminance. Motion detectors (like those found in area MT+) conceived of as spatiotemporal filters (Adelson & Bergen, 1985) may respond to brief luminance changes. A luminance transient produces equal motion energy in all direction and would thus increase the noise across the population of direction selective neurons, effectively reducing the detectability of motion transients. But contrary to this hypothesis, luminance transients in the other surface had no effect on the sensitivity to motion transients (Figure 8). Regardless of the explanation for this masking phenomenon, it occurred even in the single-task condition and thus does not change our conclusion that processing multiple features within a surface has unlimited capacity.

Next consider transient in the same feature. Same feature distractors on the other surface had the greatest influence on yes-responses (Figures 6 & 10). Whereas the masking effect discussed above may be due to input interference, the failure of selection that occurred within a feature-dimension is possibly due to output interference, or crosstalk between channels. Crosstalk and masking are distinguished by the effect of the distractor on the probability of a yes response rather than on the probability of a correct response. Crosstalk was greatest when attention was divided between-surfaces as compared to either the single-task or dual-task, within-surface conditions (Figure 8). This failure of selective attention was particularly prevalent in the luminance task—subjects responded with equal probability on trials with a single target or distractor transients. In addition, the probability of a yes-response was highest when both a target and distractor transient occurred. This suggests that some portion of the output from the channel encoding information regarding the distractor was leaking into the output of the channel carrying the cued feature information. Crosstalk was present in all conditions (Figure 6), but was most extreme when attention was divided between surfaces.

Could the failure of selective attention that we observed in our data be specific to transparent motion? The answer is probably “no.” For example, crosstalk has been reported for simultaneous binaural stimulus presentation (Gilliom & Sorkin, 1974), which, like our paradigm, presents multiple stimuli at the same time. One simple manipulation to explore would be to separate the two surfaces in space, like Vercera and Farah (1994) did with Duncan’s stimuli. Vercera and Farah found no effect of spatial separation on the dual-task deficit between objects, but whether or not we would see an effect on the level of crosstalk using our stimuli remains an open question. A second interesting manipulation that would likely effect the level of crosstalk, would be to parametrically vary the heterogeneity between target and distractor features (Lo, Howard, & Holcombe, 2012).

In our experiment, selection was worse for the luminance task then for the motion task. In the baseline conditions, sensitivity was estimated at 3.4 for motion and 3.8 for luminance. In contrast, the crosstalk gain was estimated at 0.17 for motion and 0.50 for luminance. The discrepancy for the two dimensions observed between crosstalk gains seems larger than the discrepancy observed between sensitivities. Thus, this asymmetry seems to be something to take seriously. It is possible that our chromatic feature pairs were less distinguishable than the two motion directions.

Now turn to the larger question of why might selection fail between objects. We suggest two possibilities. First, selecting two objects may result in all of the features of both objects being selected (Z. Chen & Cave, 2006; Egly, Driver, & Rafal, 1994; Yeari & Goldsmith, 2010). In this automatic selection hypothesis, object-based selection is less helpful in selecting the task relevant information in the between surface condition compared to the within surface condition. In particular, none of the irrelevant features within an object are from the same dimension as the relevant feature (e.g. motion-motion). That is not the case for the between surface condition. A second possibility is that one cannot select 2 surfaces at once and instead selects the entire stimulus. Again, this allows features from the same dimensions to interfere in the between surface condition and not in the within surface condition. These two possibilities might be distinguished by experiments in which there are 3 surfaces. Can one select 2 surfaces to prevent interference from a third?

CONCLUSION

Changes to target features within a cued surface were detected independently and without dual-task cost, consistent with an unlimited-capacity model. By contrast, when the same two target features belonged to different surfaces, detection sensitivity decreased and selection errors increased. Subjects were worse at selecting the cued feature and instead responded to changes in overall intensity, within the feature-dimension, irrespective of surface. Dividing attention across objects interferes with the ability to filter irrelevant features.

FROM PHYSIOLOGY TO PERCEPTION

Summary & conclusion

I began with this question: can attention select a single feature within an object or is selection of the whole object compulsory? Two general theories of selective attention make competing predictions. Feature-based attention posits that an observer can select a single feature, like a particular direction of motion or color. In contrast, object-based attention posits that the entire object, both its direction of motion and color, is selected (Duncan, 1984). The conflict arises when a single feature within an object is relevant to an observer. Feature-based attention predicts that attentional effects will be confined to the task-relevant feature, whereas object-based attention predicts that attentional effects will spread across the features of an attended object. My research suggests that the rules of feature-based attention do not apply when those features are bound within an object.

Differentiate between feature- and object-based selection requires a model of how attention modulates the neural response. The neural correlate of selective attention can be quantified by the response of feature-tuned neurons. According to the feature-similarity gain model—a well supported mechanistic description of feature-based attention (Boynton, 2005a; Martinez-Trujillo & Treue, 2004)—attention to a particular feature increases the response of neurons tuned to that feature while inhibiting the sensory response to irrelevant features. Evidence for object-based selection has also been shown with single-unit recordings in primates, by an increased response to both relevant and irrelevant features of an object (Fallah et al., 2007; Roelfsema et al., 1998; Wannig et al., 2007).

In humans, the effects of attention on the processing of visual stimuli can be measured non-invasively using fMRI. The hemodynamic response—measured by fMRI—in early visual cortex results from the population response of many feature-selective neurons averaged over a period of time (Boynton et al., 1996). Since our goal was to measure the response to multiple, simultaneously presented, stimulus features, disaggregating the responses to each stimulus feature from the overall fMRI time course presented a considerable methodological constraint. In order to overcome this constraint, we modulated each stimulus feature periodically through feature-space. We *frequency tagged* each stimulus component by modulating the features with unique temporal periods. By analyzing the frequency spectrum of the fMRI time course, we quantified the response to each feature via its corresponding harmonic amplitude.

Our stimulus consisted of two superimposed random dot surfaces, each defined by a unique color-motion conjunction. For every scan the subject was cued to track one of the surface's features (color or motion direction). Consistent with object-based selection, a similar attentional modulation was observed across both of the surface's features, regardless of which feature was cued. This result lead to a second testable prediction: if object-based attention selects all of a surface's features, then an observer should have no difficulty in dividing their attention across those features.

The dual-task paradigm is a common psychophysical method for measuring the effects of divided attention (Sperling & Melchner, 1978). If behavioral performance is unaffected when two tasks are performed simultaneously, then the process is said to have unlimited capacity. If object-based attention selects all of a surface's features, then there should be no cost to dividing attention across features within a surface, and perceiving changes in either feature should have unlimited capacity.

To test this hypothesis we again superimposed two dot fields; but this time the direction of motion (up/down) and the color (red/green) of each field was held constant. For each block of trials, the subject either monitored a single surface-feature, or both surface-features. Their task was to detect brief decrements in the speed or luminance (single task) or both speed and luminance (dual task) of the cued surface. Performance in the dual-task condition was equivalent to performance in the single-task condition. Consistent with object-based selection, these results suggest that an observer has unlimited capacity to detect changes in the features of an attended surface.

However, these results are also consistent with the alternative hypothesis that detecting changes in simple features, like speed and luminance, has unlimited capacity regardless of whether those features belong to the same object or not. To test this alternative, we again instructed subjects to divide their attention in order to perform the same motion and luminance detection task, except they were now asked to monitor the motion of one surface and the luminance of a second superimposed surface. Dividing attention between surfaces resulted in large costs to behavioral performance, consistent with a limited capacity model of object processing. Thus object-based attention selects all of an object's features but is limited in capacity when selecting multiple objects.

How is capacity limited when attention is divided between objects? In the most extreme case only one object could be selected at a time. If this were the case, then observers would be unable to divide their attention at all and would instead be forced to switch back-and-forth between the two objects. Because the target events occurred simultaneously in both surfaces, performance on the two tasks would be negatively correlated. Instead, we observed no correlation in performance between the two tasks, suggesting that the two tasks were performed independently. Dividing attention between objects is limited capacity, but resources are shared between the two tasks.

Surprisingly, when attention was divided between surfaces an increase in selection errors accounted for a significant portion of the reduction in dual-task performance. Selection errors were revealed by an increase in response rate conditional on same feature distractors in the wrong surface. We modeled this selective attention failure as an increase in crosstalk between the two motion/color channels. As crosstalk increases, the probability of successfully filtering a distractor signal on a given trial decreases.

Crosstalk likely depends on the level of target/distractor similarity. The more overlap in neural representation between the two, the more difficult filtering becomes. Superimposing two surfaces in space likely increases crosstalk, because targets and distractors coexist in space. But surprisingly, selection errors persisted in a pilot version of the dual-task experiment in which the two surfaces were separated in space. This result suggests that failure of selection is not an epiphenomenon of transparent motion, but an important limiting factor in divided attention. Indeed, we borrowed the term crosstalk from Navon & Miller (1987), who reported failures of selective attention in a binaural divided attention task. I can only speculate as to the ubiquity of selection errors in other divided and selective attention studies. Based on these findings, it behooves future researchers to analyze responses contingent on distractors. Crosstalk may be lurking behind the performance deficits reported in other attention studies.

In this dissertation I have provided evidence that object-based attention selects all features of a relevant object, and have shown that it is possible to simultaneously divide attention between two objects, but with limited capacity. The stimuli in both of these studies were constrained to simple features, such as motion and color, and the objects were constrained to two-dimensional surfaces. Random dot fields, along with gratings, have been widely used in vision science because their statistics are easily quantified and controlled, and because much is known about the response properties of feature-selective neurons—tuned for simple features like motion and color—found in early visual cortex. Extending this work will require a better understanding of how simple features, processed in parallel by neurons early in the visual system, are recombined later in the visual system to form the unitary perception of objects. Attention plays an important role in this process. The selective attention effects that are detailed in this dissertation are likely the tip of an iceberg of complex interactions between bottom-up sensory processing and top-down attentional modulation. With this work we move one small step closer to the ultimate goal: understanding the link between physical sensation, physiological processing, and subjective perception.

REFERENCES

Adelson, E. H., & Bergen, J. R. (1985). Spatio-temporal energy models for the Perception of Motion. *J. OPT. SOC. AM. A*, *2*(2), 1861.

Andersen, S. K., Hillyard, S. A., & Muller, M. M. (2008). Attention facilitates multiple stimulus features in parallel in human visual cortex. *Curr Biol*, *18*(13), 1006–9.

Anstis, P., & Cavanagh, P. (1983). *A minimum motion technique for judging equiluminance*. (J. D. Mollon & L. T. E. Sharpe, Eds.). London: Academic Press.

Blaser, E., Pylyshyn, Z. W., & Holcombe, A. O. (2000). Tracking an object through feature space. *Nature*, *408*(6809), 196–199. doi:10.1038/35041567

Bonnel, A.-M., & Haftser, E. R. (1998). Divided attention between simultaneous auditory and visual signals. *Perception & Psychophysics*, *60*(2), 179–190. doi:10.3758/BF03206027

Bonnel, A.-M., & Prinzmetal, W. (1998). Dividing attention between the color and the shape of objects. *Perception & Psychophysics*, *60*(1), 113–124. doi:10.3758/BF03211922

Boynton, G. M. (2005a). Attention and visual perception. *Curr Opin Neurobiol*, *15*(4), 465–9.

Boynton, G. M. (2005b). Imaging orientation selectivity: decoding conscious perception in V1. *Nat Neurosci*, *8*(5), 541–2.

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci*, *16*(13), 4207–21.

Boynton, G. M., & Finney, E. M. (2003). Orientation-specific adaptation in human visual cortex. *J Neurosci*, *23*(25), 8781–7.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. doi:10.1163/156856897X00357

Broadbent, D. E. (1958). The selective nature of learning. *Perception and communication* (pp. 244–267). Elmsford, NY, US: Pergamon Press.

Brouwer, G. J., & Heeger, D. J. (2009). Decoding and reconstructing color from responses in human visual cortex. *J Neurosci*, *29*(44), 13992–4003.

Chen, A. J.-W., Britton, M., Turner, G. R., Vytlacil, J., Thompson, T. W., & D'Esposito, M. (2012). Goal-directed attention alters the tuning of object-based representations in extrastriate cortex. *Frontiers in human neuroscience*, *6*, 187. doi:10.3389/fnhum.2012.00187

Chen, Z., & Cave, K. R. (2006). When does visual attention select all features of a distractor? *Journal of experimental psychology. Human perception and performance*, *32*(6), 1452–1464. doi:10.1037/0096-1523.32.6.1452

Davis, G., Driver, J., Pavani, F., & Shepherd, A. (2000). Reappraising the apparent costs of attending to two separate visual objects. *Vision Research*, *40*(10–12), 1323–1332. doi:10.1016/S0042-6989(99)00189-3

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*(1373), 1245 –1255. doi:10.1098/rstb.1998.0280

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, *18*(1), 193–222. doi:10.1146/annurev.ne.18.030195.001205

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, *87*(3), 272–300. doi:10.1037/0033-295X.87.3.272

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*(4), 501–517. doi:10.1037/0096-3445.113.4.501

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), 161–177.

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Attention, Perception, & Psychophysics*, *40*(4), 225–240. doi:10.3758/BF03211502

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. doi:10.1073/pnas.0611722104

Freeman, J., Brouwer, G. J., Heeger, D. J., & Merriam, E. P. (2011). Orientation Decoding Depends on Maps, Not Columns. *The Journal of Neuroscience*, *31*(13), 4792 –4804. doi:10.1523/JNEUROSCI.5160-10.2011

Gandhi, S. P., Heeger, D. J., & Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proc Natl Acad Sci U S A*, *96*(6), 3314–9.

Gardner, G. T. (1973). Evidence for independent parallel channels in tachistoscopic perception. *Cognitive Psychology*, *4*(1), 130–155. doi:10.1016/0010-0285(73)90009-1

Gilliom, J. D., & Sorkin, R. D. (1974). Sequential vs simultaneous two-channel signal detection: more evidence for a high-level interrupt theory. *The Journal of the Acoustical Society of America*, *56*(1), 157–164. doi:10.1121/1.1903247

Green, D. M., & Sweets, J. A. (1974). *Signal detection theory and psychophysics*. Huntington N.Y.: R.E. Krieger Pub. Co.

Grill-Spector, K, & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta psychologica*, *107*(1-3), 293–321.

Grill-Spector, Kalanit, & Sayres, R. (2008). Object Recognition. *Current Directions in Psychological Science*, *17*(2), 73 –79. doi:10.1111/j.1467-8721.2008.00552.x

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–30.

Haynes, J.-D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience*, *8*(5), 686–691. doi:10.1038/nn1445

Heeger, D. J., & Ress, D. (2002). What does fMRI tell us about neuronal activity? *Nat Rev Neurosci*, *3*(2), 142–51.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, N.J.: Prentice-Hall.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cogn Psychol*, *24*(2), 175–219.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nat Neurosci*, *8*(5), 679–85.

Kamitani, Y., & Tong, F. (2006). Decoding seen and attended motion directions from activity in the human visual cortex. *Curr Biol*, *16*(11), 1096–102.

Kanwisher, N., & Wojciulik, E. (2000). Visual attention: insights from brain imaging. *Nat Rev Neurosci*, *1*(2), 91–100.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased Activity in Human Visual Cortex during Directed Attention in the Absence of Visual Stimulation. *Neuron*, *22*(4), 751–761. doi:10.1016/S0896-6273(00)80734-5

Katzner, S., Busse, L., & Treue, S. (2009). Attention to the Color of a Moving Stimulus Modulates Motion-Signal Processing in Macaque Area MT: Evidence for a Unified Attentional System, *3*. doi:10.3389/neuro.06.012.2009

Lamy, D., & Egeth, H. (2002). Object-based selection: The role of attentional shifts. *Attention, Perception, & Psychophysics*, *64*(1), 52–66. doi:10.3758/BF03194557

Liu, T., Larsson, J., & Carrasco, M. (2007). Feature-Based Attention Modulates Orientation-Selective Responses in Human Visual Cortex. *Neuron*, *55*(2), 313–323. doi:10.1016/j.neuron.2007.06.030

Lo, S.-Y., Howard, C. J., & Holcombe, A. O. (2012). Feature-based attentional interference revealed in perceptual errors and lags. *Vision Research*, *63*(0), 20–33. doi:10.1016/j.visres.2012.04.021

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*(6843), 150–157. doi:10.1038/35084005

Maier, A., Wilke, M., Aura, C., Zhu, C., Ye, F. Q., & Leopold, D. A. (2008). Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. *Nature neuroscience*, *11*(10), 1193–1200. doi:10.1038/nn.2173

Mannion, D. J., McDonald, J. S., & Clifford, C. W. G. (2010). Orientation Anisotropies in Human Visual Cortex. *Journal of Neurophysiology*, *103*(6), 3465 –3471. doi:10.1152/jn.00190.2010

Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr Biol*, *14*(9), 744–51.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*(4715), 782–4.

Morgan, S. T., Hansen, J. C., & Hillyard, S. A. (1996). Selective attention to stimulus location modulates the steady-state visual evoked potential. *Proceedings of the National Academy of Sciences*, *93*(10), 4770 –4774.

Muller, M. M., Andersen, S., Trujillo, N. J., Valdes-Sosa, P., Malinowski, P., & Hillyard, S. A. (2006). Feature-selective attention enhances color signals in early visual areas of the human brain. *Proc Natl Acad Sci U S A*, *103*(38), 14250–4.

Navon, D., & Miller, J. (1987). Role of outcome conflict in dual-task interference. *Journal of Experimental Psychology. Human Perception and Performance*, *13*(3), 435–448.

Neisser, U. (1967). *Cognitive psychology*. East Norwalk, CT, US: Appleton-Century-Crofts.

O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, *401*(6753), 584–7.

O'Craven, K. M., Rosen, B. R., Kwong, K. K., Treisman, A. M., & Savoy, R. L. (1997). Voluntary Attention Modulates fMRI Activity in Human MT–MST. *Neuron*, *18*(4), 591–598. doi:10.1016/S0896-6273(00)80300-1

O'Toole, A. J., Jiang, F., Abdi, H., & Haxby, J. V. (2005). Partially Distributed Representations of Objects and Faces in Ventral Temporal Cortex. *Journal of Cognitive Neuroscience*, *17*(4), 580–590. doi:10.1162/0898929053467550

Op De Beeck, H. (2010). Against hyperacuity in brain reading: Spatial smoothing does not hurt multivariate fMRI analyses? *Neuroimage*, *49*(3), 1943–1948.

Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(2), 332–350. doi:10.1037/0096-1523.16.2.332

Palmer, J. (1995). Attention in visual search: Distinguishing four causes of a set-size effect. *Current Directions in Psychological Science*, 118–123.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*, 437–442. doi:10.1163/156856897X00366

Posner, M. I. (1980). Orienting of attention. *Q J Exp Psychol*, *32*(1), 3–25.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160–174. doi:10.1037/0096-3445.109.2.160

Regan, D. (1989). *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine*. Elsevier.

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci*, *19*(5), 1736–53.

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, *395*(6700), 376–381. doi:10.1038/26475

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nat Neurosci*, *5*(7), 631–2.

Scharff, A., & Palmer, J. (2008). Distinguishing serial and parallel models using variations of the simultaneous-sequential paradigm. *Journal of Vision*, *8*(6), 981. doi:10.1167/8.6.981

Scharff, A., Palmer, J., & Moore, C. M. (2011). Object identification has fixed capacity: Evidence for serial processing in the formation of perceptual objects. *Journal of Vision*, *11*(11), 103–103. doi:10.1167/11.11.103

Schoenfeld, M. A., Hopf, J. M., Martinez, A., Mai, H. M., Sattler, C., Gasde, A., Heinze, H. J., et al. (2007). Spatio-temporal analysis of feature-based attention. *Cereb Cortex*, *17*(10), 2468–77.

Scholl, B. J. (2001). Objects and attention: the state of the art. *Cognition*, *80*(1-2), 1–46. doi:10.1016/S0010-0277(00)00152-9

Serences, J. T., & Boynton, G. M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron*, *55*(2), 301–12.

Seymour, K., Clifford, C. W., Logothetis, N. K., & Bartels, A. (2009). The coding of color, motion, and their conjunction in the human visual cortex. *Curr Biol*, *19*(3), 177–83.

Shaw, M. (1980). Identifying Attentional and Decision-Making Components in Information Processing. *Attention and performance VIII* (pp. 277–392). Psychology Press.

Shomstein, S., & Yantis, S. (2004). Configural and contextual prioritization in object-based attention. *Psychonomic Bulletin & Review*, *11*(2), 247–253. doi:10.3758/BF03196566

Sperling, G., & Melchner, M. (1978). The attention operating characteristic: examples from visual search. *Science*, *202*(4365), 315 –318. doi:10.1126/science.694536

Sternberg, S. (1969). Memory-scanning: mental processes revealed by reaction-time experiments. *American scientist*, *57*(4), 421–457.

Toffanin, P., de Jong, R., Johnson, A., & Martens, S. (2009). Using frequency tagging to quantify attentional deployment in a visual divided attention task. *International Journal of Psychophysiology*, *72*(3), 289–298. doi:10.1016/j.ijpsycho.2009.01.006

Tootell, R. B. H., Hadjikhani, N., Hall, E. K., Marrett, S., Vanduffel, W., Vaughan, J. T., & Dale, A. M. (1998). The Retinotopy of Visual Spatial Attention. *Neuron*, *21*(6), 1409–1422. doi:10.1016/S0896-6273(00)80659-5

Tootell, R. B., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., Rosen, B. R., et al. (1995). Functional Analysis of Human MT and Related Visual Cortical Areas Using Magnetic Resonance Imaging. *The Journal of Neuroscience*, *15*(4), 3215–3230.

Townsend, J. T. (1990). Serial vs. Parallel Processing: Sometimes They Look Like Tweedledum and Tweedledee but They Can (and Should) Be Distinguished. *Psychological Science*, *1*(1), 46–54. doi:10.1111/j.1467-9280.1990.tb00067.x

Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, *12*(4), 242–248. doi:10.1080/17470216008416732

Treisman, A. M. (1964a). Verbal Cues, Language, and Meaning in Selective Attention. *The American Journal of Psychology*, *77*(2), 206–219. doi:10.2307/1420127

Treisman, A. M. (1964b). The Effect of Irrelevant Material on the Efficiency of Selective Listening. *The American Journal of Psychology*, *77*(4), 533–546. doi:10.2307/1420765

Treisman, A. M. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*(1373), 1295–1306. doi:10.1098/rstb.1998.0284

Treisman, A. M., & Gelade. (1980). A feature-integration theory of attention. *Cogn Psychol*, *12*(1), 97–136.

Treisman, A. M., Kahneman, D., & Burkell, J. (1983). Perceptual objects and the cost of filtering. *Attention, Perception, & Psychophysics*, *33*(6), 527–532. doi:10.3758/BF03202934

Treue, S., Hol, K., & Rauber, H. J. (2000). Seeing multiple directions of motion-physiology and psychophysics. *Nature neuroscience*, *3*(3), 270–276. doi:10.1038/72985

Treue, S., & Martinez-Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–9.

Valdes-Sosa, M., Cobo, A., & Pinilla, T. (1998). Transparent motion and object-based attention. *Cognition*, *66*(2), B13–B23.

Valdes-Sosa, M., Cobo, A., & Pinilla, T. (2000). Attention to object files defined by transparent motion. *J Exp Psychol Hum Percept Perform*, *26*(2), 488–505.

Vecera, S. P., & Farah, M. J. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology: General*, *123*(2), 146–160. doi:10.1037/0096-3445.123.2.146

Wannig, A., Rodríguez, V., & Freiwald, W. A. (2007). Attention to Surfaces Modulates Motion Processing in Extrastriate Area MT. *Neuron*, *54*(4), 639–651. doi:10.1016/j.neuron.2007.05.001

Watanabe, T., Harner, A. M., Miyauchi, S., Sasaki, Y., Nielsen, M., Palomo, D., & Mukai, I. (1998). Task-dependent influences of attention on the activation of human primary visual cortex. *Proceedings of the National Academy of Sciences*, *95*(19), 11489–11492. doi:10.1073/pnas.95.19.11489

Watson, S., & Kramer, A. (1999). Object-based visual selective attention and perceptual organization. *Attention, Perception, & Psychophysics*, *61*(1), 31–49. doi:10.3758/BF03211947

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, *63*(8), 1314–1329. doi:10.3758/BF03194545

Yeari, M., & Goldsmith, M. (2010). Is object-based attention mandatory? Strategic control over mode of attention. *Journal of experimental psychology. Human perception and performance*, *36*(3), 565–579. doi:10.1037/a0016897

Zeki, S., Watson, J. D., Lueck, C. J., Friston, K. J., Kennard, C., & Frackowiak, R. S. (1991). A Direct Demonstration of Functional Specialization in Human Visual Cortex. *The Journal of Neuroscience*, *11*(3), 641–649.

Zhang, W., & Luck, S. J. (2008). Feature-based attention modulates feedforward visual processing. *Nature Neuroscience*, *12*(1), 24–25. doi:10.1038/nn.2223