Dual Mechanisms for the Cross-Sensory Spread of Attention: How Much Do Learned Associations Matter?

Ian C. Fiebelkorn[1,2], John J. Foxe[1,2] and Sophie Molholm[1,2]

[1]Program in Cognitive Neuroscience, Department of Psychology, The City College of the City University of New York, NY 10031, USA and [2]The Cognitive Neurophysiology Laboratory, Program in Cognitive Neuroscience and Schizophrenia, Nathan Kline Institute for Psychiatric Research, Orangeburg, NY 10962, USA

The spread of attention-related processing across anatomically separated cortical regions plays an important role in the binding of an object's features, both within and across sensory modalities. We presented multiple exemplars of semantically congruent multisensory objects (e.g., dogs with barks) and semantically incongruent multisensory objects (e.g., guitars with barks) while recording high-density event-related potentials and tested whether highly learned associations among the multisensory features of well-known objects modulated the spread of attention from an attended visual stimulus to its paired, task-irrelevant sound. Our findings distinguish dual mechanisms for the cross-sensory spread of attention: 1) a stimulus-driven spread of attention that occurs whenever a task-irrelevant sound is simultaneously presented with an attended visual stimulus, independent of highly learned associations, and 2) a representation-driven spread of attention that occurs in response to a task-irrelevant sound that is semantically congruent with a visual target and is therefore dependent on highly learned associations. The first mechanism is thought to reflect bottom-up feature binding, whereas the second mechanism is thought to reflect the top-down activation of a stored object representation that includes the well-known object's multisensory features. When a semantically congruent, task-irrelevant sound is simultaneously presented with a well-known visual target, the combined spread of attention through both mechanisms appears additive.

Keywords: auditory, binding, electrophysiology, multisensory, visual

Introduction

Imagine a dog standing outside a screen door, barking to be let back into the house. The different features of this object—the barking dog—stimulate different types of specialized sensory receptors, and these various inputs provide complementary or redundant information that is used to identify and respond to the object. The constituent features of this or any object, whether within or across sensory modalities, are represented in anatomically separated cortical regions. A fundamental question is how these features are selected and recombined to form what is ultimately perceived as a coherent object: the so-called "binding problem."

There is substantial evidence that attention-related mechanisms play a crucial role in object-binding processes (Robertson 2003). In the influential biased competition model of visual attention, focal spatial selection results in the preferential processing of an entire object, including features that are irrelevant to the task at hand (Desimone and Duncan 1995; Duncan 2006). Indeed, experimental evidence has revealed that attention spreads both within the visual boundaries of an object—even when part of that object is outside the region of interest highlighted by spatial selection—and to the task-irrelevant visual features of that object (Egly et al. 1994; O'Craven et al. 1999; Schoenfeld et al. 2003; Wylie et al. 2004; Martinez et al. 2006, 2007; Melcher and Vidnyanszky 2006). When participants are asked to attend to an object's motion, for example, its other visual features, such as its color, also receive enhanced processing (O'Craven et al. 1999; Schoenfeld et al. 2003; Wylie et al. 2006).

Real-world objects frequently have multisensory features, and hence one would expect an effective binding mechanism to operate beyond the boundaries of visual representations. Consistent with this notion, recent studies have demonstrated that attention also spreads to a task-irrelevant sound presented with an attended visual stimulus (Busse et al. 2005; Molholm et al. 2007; Talsma et al. 2007). Molholm et al. (2007) hypothesized the existence of 2 separate processes that could lead to this object-based, cross-sensory spread of attention: 1) a "stimulus-driven" process for which no preexisting relationship between the features need be present and 2) a "representation-driven" process that involves the activation of preexisting and highly associated feature representations (in Molholm et al. 2007, the stimulus-driven and representation-driven processes are referred to as the cross-sensory spread of attention and the cross-sensory priming effect, respectively. Here we adopt new terminology for the sake of clarity). These 2 processes might be separate mechanisms through which attention binds the distributed features of an object. The stimulus-driven spread of attention is hypothesized to occur whenever a task-irrelevant sound is presented at the same time and location as an attended visual stimulus (e.g., as in Talsma et al. 2007, for arbitrarily paired tones and gratings) and is thought to reflect bottom-up feature binding. The representation-driven spread of attention, on the other hand, is hypothesized to occur in response to a task-irrelevant sound that is semantically related to a visual target (e.g., the bark of a dog when the target stimulus is the image of a dog) and is thought to reflect top-down interactions between memory and attention. In this case, the activation of one feature within the stored representation of a well-known object (e.g., a dog) leads to enhanced processing of all that object's features, including its task-irrelevant auditory features (e.g., a bark).

The representation-driven spread of attention serves as one possible mechanism through which the multisensory features of well-known objects might be treated differently than the multisensory features of novel objects. It remains to be tested, however, if the representation-driven spread of attention results from either long-term, semantically based associations among an object's multisensory features or associations established during the experimental session. It also remains to be tested if the more automatic stimulus-driven spread of

attention is influenced by highly learned associations among an object's multisensory features. Previous studies have only demonstrated the existence of this bottom-up mechanism for the cross-sensory spread of attention using novel objects (Busse et al. 2005; Talsma et al. 2007).

In the present study, we used high-density scalp-recorded event-related potentials (ERPs) to resolve these questions and to disentangle the respective contributions of the stimulus-driven and representation-driven processes to the object-based, cross-sensory spread of attention. Participants were instructed to respond to the second of consecutively presented images of a well-known object (dogs, cars, or guitars) while ignoring all sounds (i.e., a 1-back task). Unlike Molholm et al. (2007), who used a small stimulus set consisting of congruent multisensory objects (e.g., barking dogs) to examine the representation-driven spread of attention, we also included consistently paired, incongruent multisensory objects (e.g., guitars consistently paired with barks), as well as multiple exemplars of each object. The use of this expanded stimulus set allowed us to establish if the representation-driven spread of attention resulted from the activation of a preexisting, canonical representation of the well-known object, rather than experiment-specific associations. The inclusion of both object-congruent and object-incongruent multisensory trials further allowed us to establish if bottom-up feature binding, as measured through the stimulus-driven spread of attention, differs based on whether the task-irrelevant sound and its paired visual stimulus are components of the same well-known object or components of different well-known objects.

Our findings reveal dual mechanisms for the cross-sensory spread of attention with overlapping latencies and scalp topographies: 1) a stimulus-driven process that reflects an inherent bias to bind co-occurring multisensory features, regardless of whether those features are semantically congruent or incongruent and 2) a representation-driven process that requires the existence of preexisting, highly learned associations among the multisensory features. We also show that the cross-sensory spread of attention through the stimulus-driven and representation-driven processes appear additive when an image of the visual target (e.g., a dog) is paired with a congruent, task-irrelevant sound (e.g., a bark).

Materials and Methods

Subjects
Twelve neurologically normal, paid volunteers participated in the first experiment (mean age 26.7 ± 5.4 years; 4 females; 1 left handed). Data from 2 additional subjects were excluded either because of unusually poor task performance or because of bridging among the scalp electrodes. An additional 10 neurologically normal, paid volunteers participated in a second, control experiment (mean age 26.9 ± 4.3 years; 4 females; all right handed). All subjects reported normal hearing and normal or corrected-to-normal vision. The Institutional Review Boards of both the Nathan Kline Institute and the City College of the City University of New York approved the experimental procedures. Written informed consent was obtained from all subjects prior to each recording session, in line with the Declaration of Helsinki.

Stimuli
Twenty unique black and white photographs and 20 unique sounds were used to represent each of 3 well-known multisensory objects: dogs, cars, and guitars (for a total of 60 images and 60 sounds). These photographs and sounds were collected through Internet searches and later standardized. The photographs were presented for 400 ms on a 21-inch cathode-ray tube computer monitor, which was positioned 143 cm in front of participants. The centrally presented images subtended an average of 4.4° of visual angle in the vertical plane and 5.8° of visual angle in the horizontal plane and appeared on a gray background. The sounds, which were also 400 ms in duration (with 40-ms rise and fall periods), were presented at a comfortable listening level of approximately 75 dB sound pressure level over 2 JBL speakers placed at either side of the computer monitor. The photographs and sounds were presented alone and combined to form congruent pairs (e.g., dogs paired with barks) and incongruent pairs (e.g., dogs paired with the sounds of car engines), for a total of 4 stimulus types and 240 stimuli. Incongruent photographs and sounds were consistently paired throughout the experimental session: photographs of dogs were presented with sounds of car engines, photographs of cars were presented with strums of guitars, and photographs of guitars were presented with barks of dogs.

Because the inclusion of semantically incongruent audiovisual stimuli created a somewhat ecologically invalid situation where sounds might be actively suppressed, a second data set was collected after their removal. It was hypothesized that the removal of incongruent audiovisual stimuli would decrease the suppression of task-irrelevant sounds and thus increase the amplitude of the observed cross-sensory spread of attention through both the stimulus-driven and representation-driven processes. This second experiment included a total of 3 stimulus types (visual alone, auditory alone, and congruent audiovisual) and 180 stimuli.

Procedure
Participants were seated in a comfortable chair in a dimly lit and electrically shielded room and asked to keep head and eye movements to a minimum while maintaining fixation on a central cross. Eye position was monitored with horizontal and vertical electrooculogram (EOG) recordings. Visual-alone, auditory-alone, congruent audiovisual, and incongruent audiovisual stimuli were presented equiprobably and in pseudorandom order (incongruent audiovisual stimuli were excluded from the second experiment). Stimulus onset asynchrony varied randomly between 800 and 1100 ms, and a total of 180 stimuli were presented within blocks of approximately 3 min. The order of target-dog, target-car, and target-guitar blocks was pseudorandomized within a total of 36 blocks (the second experiment included a total of 27 blocks). Figure 1 shows a schematic of the experimental paradigm during a target-guitar block.

During each block, participants were asked to click the left mouse button in response to the second of consecutively presented photographs of the target object (dogs, cars, or guitars) while ignoring all sounds. For example, when dogs were the target objects, participants were instructed to make a button press response when the photograph of a dog was followed by the photograph of another (or the same) dog, regardless of intervening auditory-alone stimuli. The probability of target presentations was maintained at 6%.

Data Acquisition and Analysis
High-density continuous electroencephalography (EEG) recordings were obtained from a BioSemi ActiveTwo 168 channel system. Recordings were initially referenced online relative to a common mode active electrode and digitally sampled at 512 Hz. The continuous EEG was divided into epochs (–100 ms prestimulus to 500 ms poststimulus onset) and baseline corrected from –100 ms to stimulus onset. Trials with blinks and eye movements were automatically rejected off-line on the basis of EOG recordings. An artifact rejection criterion of ±100 μV was used at all other scalp sites to reject trials with excessive electromyography or other noise transients. To prevent contamination from motor responses associated with target detection and false alarms, all target trials (i.e., the second of consecutive trials that included an image of the target object) and false-alarm trials were discarded. Despite the removal of target trials, the use of a 1-back task allowed for the examination of responses to target objects without contamination from motor responses. EEG epochs, sorted according to stimulus type (visual alone, auditory alone, congruent audiovisual, and incongruent audiovisual) and target condition (target object and
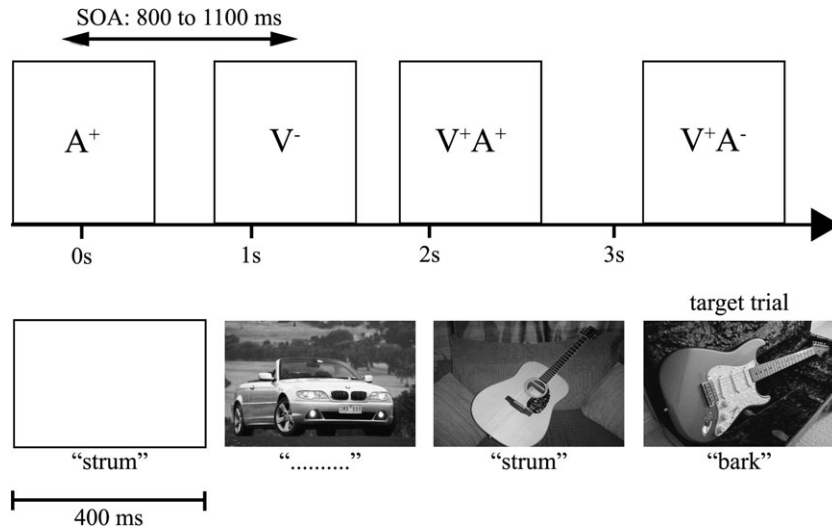
Figure 1. A schematic of the experimental design during a target-guitar block.

nontarget object), were averaged for each subject to compute ERPs. In the absence of any significant behavioral differences in response to stimuli within each of the object classes (dogs, cars, and guitars), ERPs in response to each stimulus type (4 levels) within each target condition (2 levels) were collapsed across object to increase the signal-to-noise ratio for all electrophysiological analyses. Across participants, there was an average of 355 accepted sweeps per condition. The waveforms were algebraically rereferenced to the nasion prior to data analysis. Separate group-averaged ERPs for each stimulus type within each target condition were generated for display purposes.

Statistical Analysis of Behavior

First Experiment
To determine both whether participants successfully ignored the task-irrelevant sounds and whether there were differences in behavioral performance across objects (dogs, cars, or guitars), reaction times (RTs) and error patterns were examined. Separate 2-way, repeated measures analyses of variance (ANOVAs) were conducted on RTs and hit rates (HRs), with factors of target object (3 levels: dogs, cars, or guitars) and target type (3 levels: visual alone, congruent audiovisual, or incongruent audiovisual). Pairwise comparisons were made using planned protected *t*-tests (least significant difference [LSD]). False alarms were examined to quantify the number of times each participant responded when an auditory representation of the target object was either presented alone or with an incongruent image (i.e., the image of a nontarget object). These errors, if frequent, would suggest that participants were unable to ignore task-irrelevant sounds.

Second Experiment
The analysis of behavioral results for the second experiment was identical to that from the first experiment, with the exception that there were only 2 target types (2 levels: visual alone and congruent audiovisual) instead of 3. Behavioral results were also compared across the experiments to determine whether the decreased task interference that resulted from the removal of incongruent audiovisual stimuli affected performance. To this end, between-groups ANOVAs with a factor of target type (2 levels: visual alone and congruent audiovisual) were conducted on RTs and HRs.

Statistical Analysis of Electrophysiology
Based on findings from previous studies, there were specific hypotheses about the timing and location of the ERP effects under investigation. For the first experiment, amplitude data from 7 frontocentral electrodes were averaged over a 100-ms latency window (Fig. 2). These 7 frontocentral electrodes were chosen following an
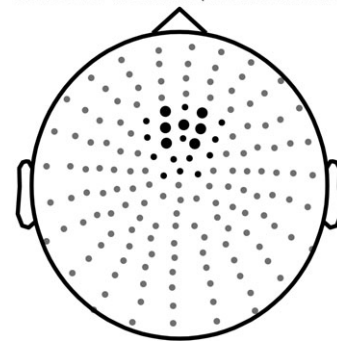
Figure 2. For all statistical analyses of ERPs, amplitude data were averaged across 7 frontocentral electrodes (large black dots). An additional 13 electrodes were included in the analysis of scalp topographies (small black dots).

examination of the grand-averaged ERP waveforms in response to task-irrelevant auditory-alone stimuli. The selected electrodes had a strong auditory N1 response, and it was reasoned that the scalp topographies for the N1 component and the attention-related processing negativity (PN) are similar (e.g., Talsma et al. 2007). A latency window from 200 to 300 ms poststimulus was chosen because the cross-sensory spread of attention from an attended visual stimulus to a simultaneously presented task-irrelevant sound, whether through the stimulus-driven or representation-driven processes, had previously been shown to onset anywhere from 180 to 250 ms poststimulus (Busse et al. 2005; Molholm et al. 2007; Talsma et al. 2007).

For the second experiment, amplitude data from the same 7 frontocentral electrodes were averaged over a shorter 50-ms latency window that was based on the timing of the cross-sensory spread of attention observed in the first experiment. This same 50-ms latency window was then used to make comparisons between the first and second experiments.

The following approaches were used to isolate the cross-sensory spread of attention-related processing that resulted from the stimulus-driven and representation-driven processes: 1) to isolate the stimulus-driven spread of attention, ERPs in response to task-irrelevant auditory representations of nontarget objects presented with attended images of nontarget objects (e.g., dogs with barks when guitars were the target objects) were compared within blocks to ERPs in response to auditory-alone representations of nontarget objects and 2) to isolate the representation-driven spread of attention, ERPs in response to task-irrelevant sounds presented with attended images of the target object

(e.g., dogs with barks when dogs were the target objects) were compared across blocks to ERPs in response to the same audiovisual stimuli when their visual component was a nontarget object (e.g., dogs with barks when guitars were the target objects) (because the behavioral task required participants to identify images as either the target object or a nontarget object, the visual stimuli were always attended. In comparison, the task-irrelevant sounds, which also included representations of both target and nontarget objects, were always unattended). Figure 3 illustrates these experimental comparisons.

First Experiment

The stimulus-driven spread of attention. A 1-way, repeated measures ANOVA, which was limited to ERPs in response to nontarget objects, was conducted with a factor of stimulus type (3 levels: auditory alone, congruent audiovisual, and incongruent audiovisual). The average amplitude (from 200 to 300 ms) of the response to an auditory-alone stimulus (A) was compared with the average amplitude of the response to an audiovisual stimulus minus the average amplitude of the response to its corresponding visual-alone stimulus (AV – V). Subtracting the visual-alone response from the audiovisual response allows for the isolation of cross-sensory effects at auditory scalp sites (e.g., Busse et al. 2005; Molholm et al. 2007). That is, the electrical signal elicited in response to the visual stimulus may volume conduct across the scalp to areas where auditory evoked potentials are typically recorded, and thus, the visual response associated with the presentation of an audiovisual stimulus could be misinterpreted as an enhanced (or reduced) auditory response.

The representation-driven spread of attention. A 2-way, repeated measures ANOVA was conducted with factors of stimulus type (3 levels: auditory alone, congruent audiovisual, and incongruent audiovisual) and target condition (2 levels: target object and nontarget object). To test for the representation-driven spread of attention in response to audiovisual stimuli, the auditory response ($A^+V^+ – V^+$ or $A^-V^+ – V^+$) to an audiovisual stimulus when its visual component was an image of the target object was compared with the auditory response ($A^-V^- – V^-$) to the same audiovisual stimulus when its visual component was an image of a nontarget object. For "congruent" audiovisual stimuli, for example, the auditory response ($A^+V^+ – V^+$) to the bark of a dog presented with the image of a dog when "dogs" were the target objects was compared with the auditory response ($A^-V^- – V^-$) to the same stimulus when either "cars" or "guitars" were the target objects. For "incongruent" audiovisual stimuli, the auditory response ($A^-V^+ – V^+$) to the sound of a car engine presented with the image of a dog when dogs were the target objects was compared with the auditory response ($A^-V^- – V^-$) to the same stimulus when guitars were the target objects.

Second Experiment

A nearly identical analysis of the electrophysiological data from the second experiment was undertaken to test for the cross-sensory spread of attention after the removal of incongruent audiovisual stimuli from the stimulus set (e.g., dogs paired with the sounds of car engines). To test for the stimulus-driven spread of attention, a 1-way, repeated measures ANOVA, which was limited to ERPs in response to nontarget objects, was conducted with a factor of stimulus type (2 levels: auditory alone and congruent audiovisual). To test for the representation-driven spread of attention, a 2-way, repeated measures ANOVA was conducted with factors of stimulus type (2 levels: auditory alone and congruent audiovisual) and target condition (2 levels: target object and nontarget object).

Electrophysiological data were also compared across experiments to determine whether the removal of incongruent audiovisual stimuli, which presumably decreased the suppression of task-irrelevant sounds, increased the amplitude of the cross-sensory spread of attention. To this end, between-groups ANOVAs were conducted on the average amplitudes of the stimulus-driven (from 246 to 296 ms) and representation-driven (from 193 to 243 ms) processes in response to congruent audiovisual stimuli. The amplitude of the cross-sensory spread of attention, whether through the stimulus-driven or the representation-driven process, was expected to increase in the absence of interference from task-irrelevant sounds (i.e., with the removal of incongruent audiovisual stimuli).

An examination of the ERP waveforms representing the stimulus-driven spread of attention from the second experiment further revealed a potential window of earlier cross-sensory integration during the time frame of the auditory N1 component at ~120 ms. A post hoc, 1-way, repeated measures ANOVA was conducted with a factor of stimulus type (auditory alone and congruent audiovisual). This analysis, which was limited to ERP responses to nontarget objects, used a 20-ms latency window that was centered on the auditory N1 component.

Both Experiments

As with the behavioral analyses, all pairwise comparisons within the electrophysiological analyses were made using planned protected *t*-tests (LSD). For all the statistical tests, both behavioral and electrophysiological, the alpha level was set at 0.05. Where appropriate, Greenhouse–Geisser corrections were made.
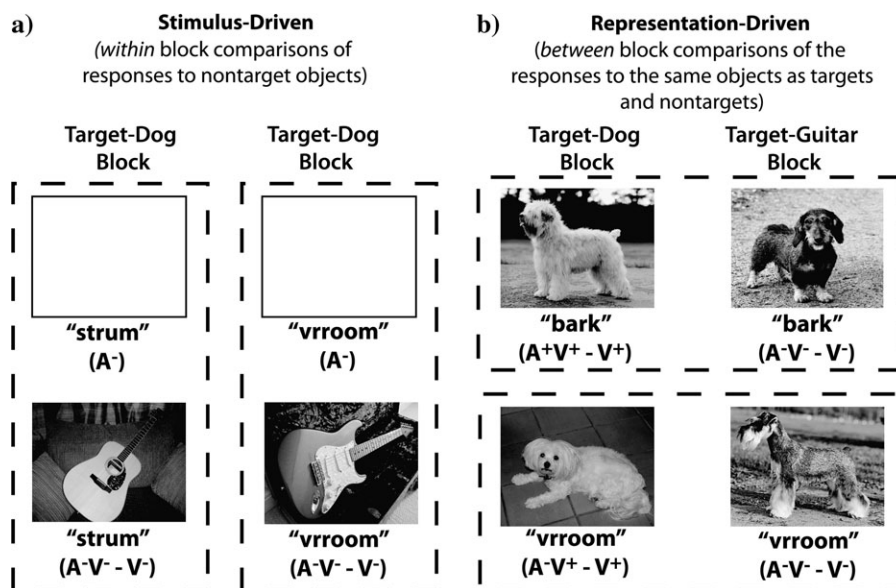
Figure 3. Examples of the experimental comparisons used to isolate the (*a*) stimulus-driven and (*b*) representation-driven processes. Responses to target and nontarget objects were collapsed across the object classes (dog, cars, and guitars).

The onset of ERP modulations was determined through running *t*-tests between amplitude measurements at consecutive time points (e.g., Molholm et al. 2002, 2007; Wylie et al. 2003). Onset was considered to be the first of 10 consecutive, statistically significant time points. This criterion was chosen because the likelihood of getting 10 false positives in a row is extremely low (Guthrie and Buchwald 1991). The use of running *t*-tests with a strict criterion controls for the type I errors that result from multiple comparisons.

Scalp Topographies

To test whether attention-related processing from the stimulus-driven and representation-driven processes was attributable to different intracranial current sources, scalp topographies of the significant attention-related effects were compared. These comparisons were made using topography-normalized voltages (based on the amplitudes of the difference waves) from 20 frontocentral electrodes (Fig. 2) (McCarthy and Wood 1985). A 2-way, repeated measures ANOVA was conducted with factors of cross-sensory attention effect (4 levels: a stimulus-driven spread of attention in response to congruent audiovisual stimuli, a stimulus-driven spread of attention in response to incongruent audiovisual stimuli, a representation-driven spread of attention in response to congruent audiovisual stimuli, and a representation-driven spread of attention in response to auditory-alone stimuli) and anterior–posterior position (4 levels: frontal, frontocentral, central, and posterior). Significant interactions between the attention-related effects and the electrode position factor would indicate significant differences among the scalp topographies and thus differences among the intracranial current sources (e.g., Talsma and Kok 2001; Talsma et al. 2007).

Results

Behavioral Results

First Experiment

As shown in Table 1, RTs when images of guitars were the target were somewhat slower than RTs when images of either dogs or cars were the target (~14 ms slower). RTs were also somewhat slower in response to congruent audiovisual targets (i.e., when the visual target was paired with a semantically congruent, task-irrelevant sound) relative to RTs in response to either incongruent audiovisual targets or visual-alone targets (~8 ms slower). A repeated measures ANOVA with factors of target objects (3 levels: dogs, cars, and guitars) and target types (3 levels: visual alone, congruent audiovisual, and incongruent audiovisual) revealed a significant main effect of target objects ($F_{2,22} = 9.0$, $P = 0.001$), whereas the main effect of target types did not reach significance ($F_{2,22} = 2.7$, $P = 0.09$). There were also no significant interactions between target objects and target types ($F_{4,44} = 0.6$, $P = 0.95$). Pairwise comparisons across the target objects revealed that RTs when images of guitars were the target were indeed significantly slower than RTs when images of either dogs ($P = 0.002$) or cars ($P = 0.006$) were the target.

This difference might reflect the fact that participants were less familiar with guitars than dogs or cars.

Molholm et al. (2004), who asked participants to simultaneously detect visual and auditory representations of a specific well-known animal (e.g., a dog), reported that RTs in response to congruent audiovisual targets were faster than those in response to both incongruent audiovisual targets (e.g., a barking cat when cats were the target object) and visual-alone targets. The absence of this trend in the present data was interpreted as evidence that participants had successfully ignored the task-irrelevant sounds. This interpretation was further supported by the negligible number of false alarms in response to auditory-alone representations (e.g., a bark) of the visual target (e.g., the image of a dog) or false alarms in response to auditory representations of the visual target paired with incongruent images of nontargets (e.g., the image of a guitar). Across all 12 participants, less than 1% of responses were categorized as false alarms, and only 3 false alarms fit the above description.

The HR collapsed across target objects and target types was 93% and varied little within or across these factors (Table 1). A repeated measures ANOVA revealed no significant main effects or interactions, indicating that task performance was approximately equivalent across target objects ($F_{2,22} = 0.2$, $P = 0.81$) and target types ($F_{2,22} = 0.1$, $P = 0.89$). The absence of significant differences among HRs, as well as the similarity among RTs, served as justification to collapse the ERPs in response to each stimulus type (auditory alone, visual alone, congruent audiovisual, and incongruent audiovisual) across the object classes (dogs, cars, and guitars), which increased the signal-to-noise ratio for all electrophysiological analyses.

Second Experiment

The behavioral results from the second experiment, which did not include incongruent audiovisual stimuli (e.g., dogs paired with sounds of car engines), revealed no significant main effects or interactions for RTs or HRs. That is, both RTs and HRs were approximately equivalent across target objects ($F_{2,14} = 0.09$, $P = 0.92$; $F_{2,14} = 1.32$, $P = 0.30$) and target types ($F_{1,7} = 0.10$, $P = 0.78$; $F_{1,7} = 0.22$, $P = 0.66$). Similar to the results from the first experiment, less than 1% of the responses across all 10 participants were categorized as false alarms, and only 2 false alarms occurred in response to an auditory-alone representation of the visual target.

Table 2 compares the behavioral results from the first and second experiments. It was reasoned that the inclusion of incongruent audiovisual stimuli in the first experiment would create a somewhat ecologically invalid situation where task-irrelevant sounds might be actively suppressed to prevent them from interfering with the behavioral task. In support of this notion, performance improved when task interference from task-irrelevant sounds decreased (i.e., when incongruent audiovisual stimuli were removed from the stimulus set).

Table 1
Average RTs and percent hits sorted by target objects and target types

	Dogs			Cars			Guitars		
	V^+	V^+A^+	V^+A^-	V^+	V^+A^+	V^+A^-	V^+	V^+A^+	V^+A^-
RTs	452(71)	461(66)	452(69)	454(70)	463(59)	456(71)	468(72)	475(65)	469(78)
%Hit	93(7)	94(8)	94(6)	93(5)	93(8)	94(6)	94(7)	93(7)	92(7)

Note: Visual alone (V^+), congruent audiovisual (V^+A^+), incongruent audiovisual (V^+A^-). All times in milliseconds. Standard deviations in parentheses.

Table 2
Average RTs and percent hits sorted by experiment and target types

	First experiment		Second experiment	
	V^+	V^+A^+	V^+	V^+A^+
RTs	457(69)	461(65)	398(40)	398(43)
%Hits	93(6)	93(7)	97(2)	97(2)

Note: Visual alone (V^+) and congruent audiovisual (V^+A^+). All times in milliseconds. Standard deviations in parentheses.

Between-groups ANOVAs revealed that RTs from the second experiment, relative to those from the first experiment, were significantly faster in response to both visual-alone targets ($F_{1,20} = 4.80$, $P = 0.04$) and congruent audiovisual targets ($F_{1,20} = 7.04$, $P = 0.02$). In addition to these faster RTs, HRs from the second experiment were also higher in response to both visual-alone targets and congruent audiovisual targets. Between-groups ANOVAs, however, revealed that HR differences across the experiments did not reach statistical significance ($F_{1,20} = 2.81$, $P = 0.11$; $F_{1,20} = 3.24$, $P = 0.09$).

Electrophysiological Results

First Experiment

The stimulus-driven spread of attention. The ERP waveforms shown in Figure 4 illustrate the stimulus-driven spread of attention that occurs whenever a task-irrelevant sound is paired with an attended visual stimulus. A repeated measures ANOVA with a single factor of stimulus type (auditory alone, congruent audiovisual, and incongruent audiovisual), which was limited to ERPs in response to nontarget objects, revealed a significant main effect ($F_{2,22} = 7.1$, $P = 0.01$). Pairwise comparisons across the stimulus types further revealed that ERPs in response (from 200 to 300 ms) to congruent and
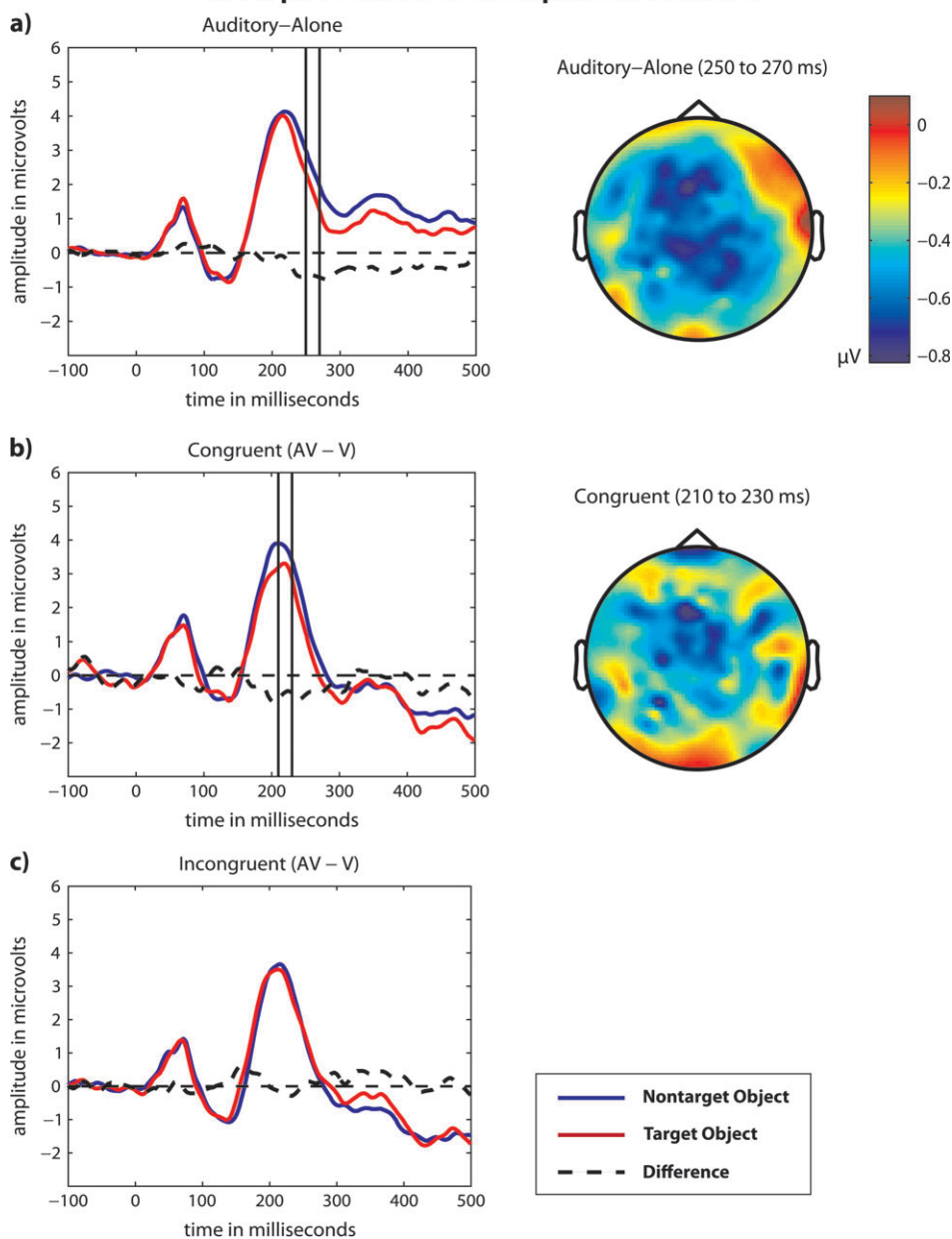
incongruent audiovisual stimuli (AV – V) were significantly more negative than ERPs in response to auditory-alone stimuli ($P = 0.03$, $P = 0.01$). In comparison, the ERPs in response to congruent and incongruent audiovisual stimuli (AV – V, limited to ERPs in response to nontarget objects) did not differ significantly ($P = 0.48$). Running t-tests revealed that the onset of the stimulus-driven spread of attention in response to congruent and incongruent audiovisual stimuli was also approximately equivalent, starting at 246 and 236 ms, respectively. These data indicate that the stimulus-driven spread of attention from an attended visual stimulus to its paired, task-irrelevant sound is independent of highly learned associations among a well-known object's multisensory features.

The representation-driven spread of attention. The ERP waveforms shown in Figure 5 illustrate the representation-driven spread of attention, hypothesized to result from the activation of a cortical representation established through a lifetime of experiences with an object. A repeated measures ANOVA with factors of stimulus type (3 levels: auditory alone, congruent audiovisual, and incongruent audiovisual) and target condition (2 levels: target object and nontarget object) revealed both significant main effects ($F_{2,22} = 4.8$, $P = 0.02$; $F_{1,11} = 5.6$, $P = 0.04$) and a significant interaction ($F_{2,22} = 4.2$,
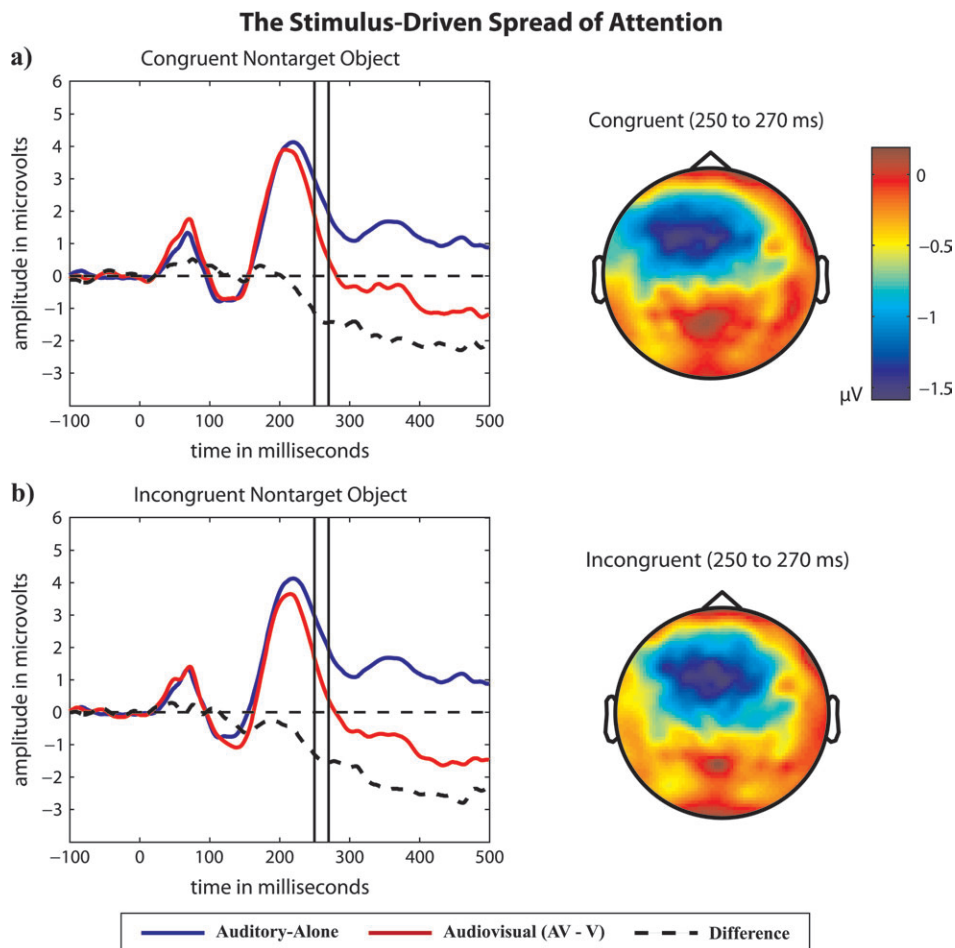
Figure 4. The stimulus-driven spread of attention from an attended visual stimulus to a paired (*a*) congruent or (*b*) incongruent, task-irrelevant sound. In both cases, ERPs were limited to those in response to nontarget objects. Flattened voltage maps were derived from ERP difference waves.

The Representation-Driven Spread of Attention

Figure 5. The representation-driven spread of attention that occurs in response to a task-irrelevant sound that is semantically congruent with a visual target. ERP waveforms are shown for (*a*) auditory-alone, (*b*) congruent audiovisual, and (*c*) incongruent audiovisual stimuli. Flattened voltage maps, derived from the ERP difference waves, are shown for (*a*) auditory-alone and (*b*) congruent audiovisual stimuli.

$P = 0.02$). Pairwise comparisons between target conditions (2 levels: target object and nontarget object) within the stimulus types revealed that ERPs in response to auditory-alone stimuli ($P = 0.04$) and ERPs (AV – V) in response to congruent audiovisual ($P = 0.04$) stimuli were significantly more negative when those stimuli were representations of the visual target than when those stimuli were representations of visual nontargets. However, ERPs (AV – V) in response to incongruent audiovisual stimuli did not significantly differ across the target conditions ($P = 0.91$). These data indicate that the representation-driven spread of attention depends on preexisting, highly learned associations among a well-known object's multisensory features. As in Molholm et al. (2007), the

representation-driven spread of attention occurred somewhat earlier when task-irrelevant sounds were paired with a visual target, starting at 193 ms compared with 236 ms when task-irrelevant sounds were presented alone.

Scalp topographies. The voltage maps associated with the cross-sensory spread of attention that results from both the stimulus-driven and representation-driven processes (see Figs 4 and 5) resembled the distribution of the late frontal negativity typically observed in response to an attended sound (a PN or an Nd component; Näätänen et al. 1978; Hansen and Hillyard 1980). Previous studies have reported similar scalp topographies corresponding to the stimulus-driven spread of attention from
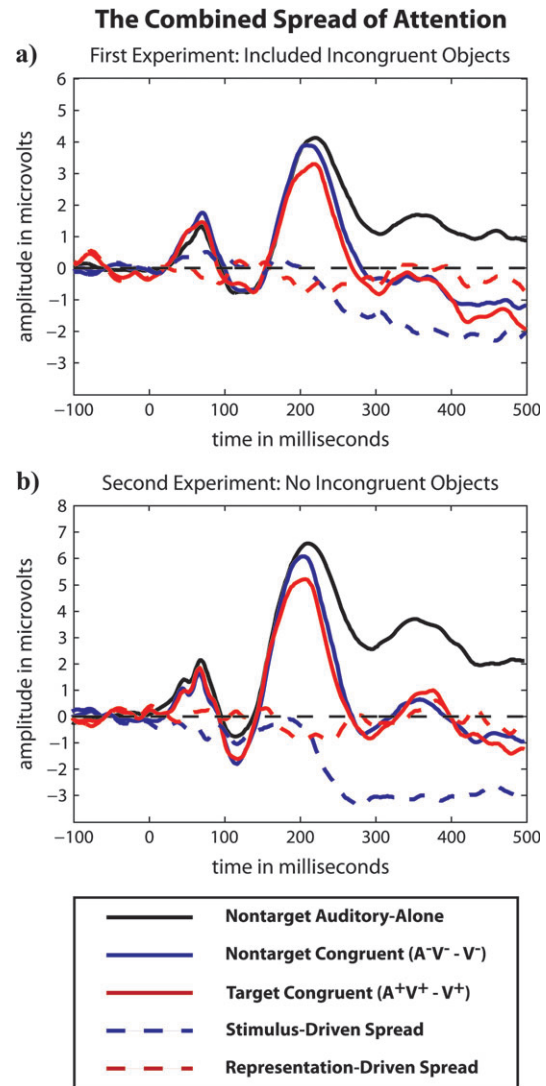
an attended visual stimulus to its paired, task-irrelevant sound (Busse et al. 2005; Talsma et al. 2007) and corresponding to the representation-driven spread of attention from a visual target to a semantically congruent, task-irrelevant sound (Molholm et al. 2007). For the data presented here, a comparison of the scalp topographies was conducted to test whether different underlying current sources generated the stimulus-driven and representation-driven processes. For this analysis, significant interactions between the attention-related effects and electrode positions would indicate significant differences among the scalp topographies; however, a 2-way, repeated measures ANOVA conducted with attention effects (4 levels: a stimulus-driven spread of attention in response to congruent audiovisual stimuli, a stimulus-driven spread of attention in response to incongruent audiovisual stimuli, a representation-driven spread of attention in response to congruent audiovisual stimuli, and a representation-driven spread of attention in response to auditory-alone stimuli) and anterior–posterior position (4 levels: frontal, frontocentral, central, and posterior) as factors revealed no significant interactions ($F_{9,99} = 1.9$, $P = 0.12$).

Second Experiment

Repeated measures ANOVAs revealed that both mechanisms leading to the cross-sensory spread of attention remained significant when incongruent audiovisual stimuli were removed from the stimulus set. For the stimulus-driven spread of attention, a within blocks analysis limited to stimuli representing nontarget objects revealed that ERPs in response (from 246 to 296 ms) to congruent audiovisual stimuli (AV – V) were significantly more negative than ERPs in response to auditory-alone stimuli ($F_{1,9} = 11.6$, $P = 0.01$). For the representation-driven spread of attention, a between-blocks analysis that included all stimuli (i.e., stimuli representing either target or nontarget objects) revealed that ERPs in response (from 236 to 286 ms) to auditory-alone representations of the visual target and ERPs in response (from 193 to 243 ms) to congruent audiovisual targets (AV – V) were both significantly more negative than ERPs in response to the same stimuli when those stimuli were representations of visual nontargets ($F_{1,9} = 12.9$, $P = 0.01$).

The ERP waveforms shown in Figure 6 illustrate the combined, cross-sensory spread of attention (i.e., from both attention-related mechanisms) in response to a congruent audiovisual stimulus when its visual component was a target. A comparison between the electrophysiological results from the first and second experiments, which used the shorter 50-ms latency windows, indicated that the presence of incongruent audiovisual stimuli diminished the stimulus-driven spread of attention from an attended visual stimulus to its paired, task-irrelevant sound ($F_{1,20} = 4.3$, $P = 0.05$). Counter to our expectations, the presence of incongruent audiovisual stimuli did not diminish the representation-driven spread of attention ($F_{1,20} = 0.0$, $P = 0.87$).

An examination of the ERP waveforms displayed in Figure 6b (recorded when incongruent audiovisual stimuli were removed from the stimulus set) also suggests the possible emergence of an earlier window of cross-sensory integration during the time frame of the auditory N1 component (at ~120 ms). A post hoc, 1-way, repeated measures ANOVA (limited to nontarget objects) with a single factor of stimulus type (auditory alone and congruent audiovisual), however, only approached statistical significance ($F_{1,9} = 4.2$, $P = 0.07$).

The Combined Spread of Attention

a) First Experiment: Included Incongruent Objects

b) Second Experiment: No Incongruent Objects

Legend:
- Nontarget Auditory-Alone
- Nontarget Congruent (A⁻V⁻ - V⁻)
- Target Congruent (A⁺V⁺ - V⁺)
- Stimulus-Driven Spread
- Representation-Driven Spread

Figure 6. The combined cross-sensory spread of attention from the (*a*) first and (*b*) second experiments.

Discussion

Research from the visual domain strongly suggests an inherent bias to process objects as wholes, even when attention is focused on constituent features (e.g., Stroop 1935; Egly et al. 1994; O'Craven et al. 1999; Blaser et al. 2000; Schoenfeld et al. 2003; Martinez et al. 2006). Both brain imaging and electrophysiological studies have demonstrated an object-based spread of attention to feature representations in anatomically distributed cortical regions (e.g., O'Craven et al. 1999; Schoenfeld et al. 2003). These studies collectively suggest that attention-related processing serves as a mechanism through which visual features are bound to form the perception of a coherent object (e.g., Robertson 2003; Desimone and Duncan 1995; Duncan 2006; Serences and Yantis 2006). Although such models for feature binding are typically discussed in the context of an object's visual features, there is no obvious reason why the binding of an object's multisensory features would not similarly occur through attentional mechanisms. Indeed, recent research suggests that the whole-object bias, which seems to operate through the spread of attention, extends across

sensory boundaries (Busse et al. 2005; Molholm et al. 2007; Talsma et al. 2007).

Here we used electrophysiological measures of selective attention to investigate the cross-sensory spread of attention from an attended visual stimulus to a task-irrelevant sound, examining the interplay between bottom-up and top-down attentional processes. More specifically, we investigated whether highly learned associations among the multisensory features of a well-known object, formed through a lifetime of experiences with that object, modulate the spread of attention that contributes to the binding of its multisensory features.

Our findings distinguish dual mechanisms for the object-based cross-sensory spread of attention: 1) a bottom-up, stimulus-driven spread of attention that seems to be unaffected by highly learned associations among the multisensory features of a well-known object and 2) a top-down, representation-driven spread of attention that is dependent on highly learned associations among the multisensory features of a well-known object. We propose that the stimulus-driven spread of attention reflects an inherent bias to process objects as wholes, with its induction based entirely on the temporal and spatial properties of a multisensory object's features. The representation-driven spread of attention, on the other hand, only occurs in response to sounds that are semantically related to a visual target, reflecting the activation of a preexisting cortical representation of a well-known object that includes the object's multisensory features. We also find that when a semantically congruent, task-irrelevant sound is simultaneously presented with a visual target, the effects of the stimulus-driven and representation-driven processes appear to be additive (see Fig. 6).

The Representation-Driven Spread of Attention

Molholm et al. (2007) used well-known multisensory objects to demonstrate a cross-sensory spread of attention that occurred only when a task-irrelevant sound was semantically congruent with a visual target. It was hypothesized that this spread of attention occurred because selectively attending for an object in the visual modality led to the activation of a preexisting object representation that included the highly associated multisensory features of the visual target. The first goal of the present study was to establish whether this representation-driven spread of attention in fact results from long-term, semantically based associations among an object's multisensory feature representations or instead from associations formed as a result of repeated pairings during the experimental session, as Molholm et al. (2007) used only a single exemplar for each of the repeatedly presented multisensory objects.

We devised the following approach to limit the formation of associations between specific audiovisual features during the experimental session: 1) multiple exemplars were used within each object class, so participants were forced to use a conceptual definition of the visual target rather than rely on the repetition of specific features (lines and curves at specific locations) and 2) sounds were task irrelevant throughout the experimental session to avoid task carryover effects. Despite these controls, there was significant attention-related processing of task-irrelevant sounds when those sounds were semantically congruent with the visual target, both when the sounds were presented with the visual target and when the sounds were presented alone. The absence of an attentional spread to semantically incongruent, task-irrelevant sounds

when paired with a visual target, despite the consistent pairing of mismatched sounds and images on incongruent trials (e.g., the sound of a car engine was consistently paired with the image of a dog), further demonstrates that experiment-specific associations do not account for the representation-driven spread of attention.

The above findings point to a process whereby a lifetime of experiences with an object (e.g., a dog) establishes strong associations among its cortically distributed multisensory representations, such that the activation of its visual feature representation leads to the activation of a distributed network of feature representations. In other words, the sustained activation of a well-known object's visual feature representation also creates a processing advantage for incoming sensory signals that match that object's highly associated nonvisual feature representations. This biasing toward the constituent features of a well-known object, including those features that are irrelevant to the task at hand, could conceivably contribute both to binding multisensory objects and to resolving competitions among the multitude of objects simultaneously present in a real-world environment (Desimone and Duncan 1995; Duncan 2006).

It is of note that the representation-driven spread of attention that occurs when a semantically congruent, task-irrelevant sound is simultaneously presented with a visual target does not elicit a "sustained" response as seen for the representation-driven spread of attention in response to an auditory-alone stimulus and the stimulus-driven spread of attention. It might therefore be argued that this ERP modulation instead reflects a non–attention-related cognitive process. Based on the following rationale, however, we conclude that the representation-driven effect is indeed associated with attention-related processing: 1) an apparent reduction in the amplitude of the auditory P2 component is a constituent of the PN and a well-documented marker for attention-related processing in general (e.g., Näätänen et al. 1978; Hansen and Hillyard 1980; Crowley and Colrain 2004), 2) attention effects are not necessarily sustained (e.g., Hansen and Hillyard 1980), and 3) known electrophysiological markers of cognitive processes that are sensitive to the semantic content of the stimulus onset at ~300 ms at the earliest (the class of so-called "N400" effects: e.g., Kutas and Hillyard 1980; Molholm et al. 2004). To elaborate on this second point, a more negative ERP (at ~300 ms or later) has previously been associated with the processing of semantically incongruent stimuli, relative to the processing of semantically congruent stimuli. In the present findings, it is instead the semantically congruent stimuli that evoke a more negative response (Fig. 5b). One possibility that arises from examination of the waveforms in this condition is that the attention effect does not appear to be sustained because there is an additional process superimposed in the later time frame, perhaps a process related to multisensory integration.

The Stimulus-Driven Spread of Attention

Previous research has also demonstrated a nontarget-specific, cross-sensory spread of attention that occurs whenever a task-irrelevant sound is simultaneously presented with an attended visual stimulus (Busse et al. 2005; Talsma et al. 2007). This stimulus-driven spread of attention has been hypothesized to reflect bottom-up feature binding. Previous research,

however, has not addressed whether the multisensory features of frequently encountered objects become "hard-wired," such that this bottom-up attentional spread is facilitated for well-known multisensory objects. Another goal of the present study was thus to determine whether the stimulus-driven spread of attention is modulated by highly learned feature associations.

To this end, we compared the stimulus-driven spread of attention from an attended visual stimulus with its paired task-irrelevant sound when that sound was either object congruent (e.g., barks paired with dog images) or object incongruent (e.g., the sounds of car engines paired with dog images). If highly learned associations affect bottom-up feature binding through the stimulus-driven spread of attention, we would expect to observe enhanced processing when an attended visual stimulus is paired with a semantically congruent, task-irrelevant sound, relative to when the same visual stimulus is paired with a semantically incongruent, task-irrelevant sound. On the contrary, our findings indicate that the stimulus-driven spread of attention is independent of highly learned associations among multisensory features: the stimulus-driven spread of attention from the image of a dog to the congruent, task-irrelevant bark of a dog, for example, was equivalent to the stimulus-driven spread of attention from the image of a dog to the incongruent, task-irrelevant sound of a car engine (Fig. 5). This leads us to conclude that the brain's default mode is to bind multisensory features based on their temporal and spatial properties, without regard for the learned associations (or lack thereof) among those features. This accords well with subjective experience, where a barking cat would be odd, but the multisensory features would still be perceived as a coherent object.

The Combined Cross-Sensory Spread of Attention

The data presented here demonstrate that the stimulus-driven and representation-driven processes are separate mechanisms through which attention spreads across sensory modalities to encompass an object's task-irrelevant nonvisual features. The first line of evidence is that the stimulus-driven spread of attention occurs in response to both semantically congruent and semantically incongruent multisensory objects, whereas the representation-driven spread of attention occurs only in response to semantically congruent multisensory objects. Further evidence that these are distinct processes is gained from another defining characteristic: unlike the stimulus-driven spread of attention, the representation-driven spread of attention does not require the co-occurrence of an attended visual stimulus.

Despite being separate mechanisms, the stimulus-driven and representation-driven processes seem to have similar outcomes: both processes lead to a cross-sensory spread of attention with overlapping latencies and scalp topographies. Visual examination of the scalp topographies associated with stimulus-driven and representation-driven processes does reveal a somewhat more anterior frontocentral distribution for the stimulus-driven attentional spread (Figs 4 and 5), however, suggesting the possibility of different underlying neural generators. But these differences between the scalp distributions were not statistically significant, and another possibility is that the smaller effect size associated with the representation-driven spread of attention simply yielded a noisier distribution. It is also of note that in Molholm et al. (2007), where the representation-driven spread

of attention was of greater amplitude, the corresponding scalp distribution appeared to be somewhat more anterior (see Figure 3 in Molholm et al. 2007), in line with the more anterior distribution of the stimulus-driven spread of attention seen here and elsewhere (Busse et al. 2005; Talsma et al. 2007). Nevertheless, such eyeball comparisons are far from definitive, and neuroimaging studies would be much better suited to detecting subtle differences in the underlying neural generators of these 2 processes.

Because the inclusion of semantically incongruent multisensory objects in the stimulus set created a somewhat ecologically invalid situation where sounds might be actively suppressed, we collected a second data set after the removal of incongruents from the stimulus set. This second data set was meant to test whether the inclusion of semantically incongruent multisensory objects in the present study had led to the smaller effect sizes associated with the representation-driven spread of attention relative to the greater effect sizes observed by Molholm et al. (2007), where only semantically congruent multisensory objects were presented. Whereas the removal of incongruents from the stimulus set increased the amplitude of the stimulus-driven spread of attention, the amplitude of the representation-driven spread of attention was statistically equivalent across the 2 data sets collected here (Fig. 6). Instead, it might have been the inclusion of multiple exemplars of each well-known object in the present study that led to the smaller amplitude representation-driven spread of attention. The inclusion of multiple exemplars required participants to decode a given stimulus as a size-invariant, position-invariant member of an object class (e.g., a dog), rather than as a specific object (e.g., the dog). Molholm et al. (2007), on the other hand, used a single exemplar to represent each well-known object. The repetition of a single exemplar, which constrained the representation set that needed to be matched to determine the presence or absence of a target on each trial, might have facilitated the decoding of the task-irrelevant sound's identity. This easier and more consistent decoding of the task-irrelevant sound might have led to a larger amplitude representation-driven spread of attention.

A Model for the Cross-Sensory Spread of Attention and Concluding Remarks

Our findings suggest a model of attentional spread in which object-based selection following spatial selection leads to the stimulus-driven spread of attention, independent of semantic processing: 1) spatial selection determines the relevant location for further processing (Treisman and Gelade 1980), 2) attention spreads within the visual boundaries of the object (object-based selection; Egly et al. 1994; Martinez et al. 2006, 2007), and 3) attention spreads to coincident multisensory features that fall within the visual boundaries established through object-based selection. This model for the stimulus-driven spread of attention from an attended visual stimulus to its task-irrelevant multisensory features is based on both the timing of the associated ERP effects and the apparent dominance of vision in feature integration and object recognition (e.g., Greene et al. 2001; James et al. 2002; Molholm et al. 2004, 2007; Amedi et al. 2005; Talsma et al. 2007). First, our results, as well as those of previous studies, show that the stimulus-driven spread of attention from an attended visual stimulus to its paired task-irrelevant sound onsets between 220 and 250 ms poststimulus: a time frame that

1) follows the spread of attention within the visual boundaries of an object, which has been shown to onset at ~160 ms poststimulus (e.g., Martinez et al. 2006), and 2) is similar to the time frame for the spread of attention to task-irrelevant visual features, which has been shown to also onset between 220 and 250 ms poststimulus (e.g., Schoenfeld et al. 2003). Second, several previous studies have shown that the stimulus-driven spread of attention from an attended sound to its paired task-irrelevant visual stimulus is either significantly reduced in magnitude relative to the visual-to-auditory spread of attention or altogether nonexistent (e.g., Teder-Sälejärvi et al. 1999; Molholm et al. 2007; Talsma et al. 2007). Talsma et al. (2007), for example, showed no stimulus-driven spread of attention from an attended tone to its paired, task-irrelevant horizontal grating (i.e., during an attend-auditory condition) but did show a significant stimulus-driven spread of attention from an attended horizontal grating to its paired, task-irrelevant tone (i.e., during an attend-visual condition).

The model presented here posits that induction of the stimulus-driven spread of attention occurs whenever a task-irrelevant sound overlaps an attended visual stimulus in time and space, but previous findings might seem to challenge this assertion. Van der Burg et al. (2008), for example, have demonstrated that a spatially noninformative auditory stimulus guides attention toward a synchronized visual event within a cluttered, continuously changing visual environment. The spatially noninformative auditory stimulus increases the salience of the visual target, leading to attentional capture. This shows that some multisensory integration effects might not be constrained by space. Data from our laboratory demonstrated that auditory–somatosensory interactions starting at approximately 50 ms poststimulus were not affected by the spatial alignment of the unisensory signals (Murray et al. 2005). However, in comparison to the later multisensory effects described in the present study (i.e., the spread of attentional processing from an attended visual stimulus to a task-irrelevant sound), these earliest multisensory integration effects, which may operate through direct connections between primary sensory cortices, seem likely to contribute to signal detection rather than directly contributing to higher order multisensory object processing. Busse et al. (2005), on the other hand, made clever use of the ventriloquism illusion, whereby perceptual localization of a sound is captured by a simultaneously presented but spatially separated visual stimulus, to attribute the stimulus-driven spread of attention to an object-based late selection process (an ERP modulation starting ca., 220 ms poststimulus). This spread of attention-related processing across modalities and space, occurring during the time frame of the effects we describe in the present study, may initially seem incompatible with our model for the stimulus-driven spread of attention, which states that attention spreads to task-irrelevant multisensory features that fall within the visual boundaries established through object-based selection. That is, Busse et al. (2005) did not present the task-irrelevant sound at the same physical location as the attended visual stimulus. However, induction of the ventriloquism illusion, whereby the localization of the sound was drawn toward the location of the visual stimulus, indicates that subjectively the task-irrelevant sound did fall within the visual boundaries of the object. According to this interpretation, the findings of Busse et al. (2005) fit well within our model for the stimulus-driven spread of attention.

Unlike the stimulus-driven spread of attention, the representation-driven spread of attention relies on the semantic properties of the stimulus. The activation of a well-known object's visual representation leads to the activation of its highly associated auditory representations, and a stimulus that partially matches an activated auditory representation will receive enhanced processing even when it is presented alone (i.e., without a semantically congruent visual target). Because the representation-driven spread of attention occurs in response to an auditory-alone stimulus, it cannot be strictly considered a vehicle for binding the simultaneously presented multisensory features of an object into a coherent whole. But when a semantically congruent, task-irrelevant sound (e.g., a bark) is paired with a visual target (e.g., the image of a dog when dogs are the visual target), the cross-sensory spread of attention that results from the stimulus-driven and representation-driven processes appears to be additive (Fig. 6), suggesting that the representation-driven spread of attention can contribute to multisensory feature binding.

Our model implies that object processing (object-based selection) leads to the stimulus-driven spread of attention, following the selection of a relevant location (spatial selection). An alternative model, based entirely on the supramodal properties of spatial attention, would posit that spatial selection leads to the stimulus-driven spread of attention independent of object-based selection (Hillyard et al. 1984; Eimer and Shröger 1998; Teder-Sälejärvi et al. 1999; McDonald et al. 2000, 2003). In contrast to the stimulus-driven spread of attention, we hypothesize that the top-down, representation-driven spread of attention operates independently of spatial selection. In this case, a task-irrelevant sound that is semantically congruent with the visual target will receive enhanced processing even when it is presented at an unattended location. (For examples of visual features receiving enhanced processing outside the spotlight of spatial selection, see the following: Chelazzi et al. 1993; Treue and Martinez-Trujillo 1996; Chelazzi et al. 1998; Saenz et al. 2003.) Such findings would indicate that induction of the representation-driven spread of attention, antipodal to the stimulus-driven spread of attention, is relatively independent of a multisensory object's temporal and spatial properties but dependent on its semantic properties. The resolution of these remaining issues will further expand our understanding of the interplay between attentional deployment and object processing that leads to multisensory feature binding.

References

Amedi A, Von KK, Van Atteveldt NM, Beauchamp MS, Naumer MJ. 2005. Functional imaging of human crossmodal identification and object recognition. Exp Brain Res. 166:559–571.

Blaser E, Pylyshyn W, Holcombe A. 2000. Tracking an object through feature space. Nature. 408:196–199.

Busse L, Roberts KC, Crist RE, Weissman DH, Woldorff MG. 2005. The spread of attention across modalities and space in a multisensory object. Proc Natl Acad Sci USA. 102:18751–18756.

Chelazzi L, Duncan J, Miller EK, Desimone R. 1998. Responses of neurons in inferior temporal cortex during memory-guided visual search. J Neurophysiol. 80:2918–2940.

Chelazzi L, Miller EK, Duncan J, Desimone R. 1993. A neural basis for visual search in inferior temporal cortex. Nature. 363:345–347.

Crowley KE, Colrain IM. 2004. A review of the evidence for P2 being and independent component process: age, sleep and modality. Clin Neurophysiol. 115:732–744.

Desimone R, Duncan J. 1995. Neural mechanisms of selective visual attention. Annu Rev Neurosci. 18:193–222.

Duncan J. 2006. EPS Mid-Career Award 2004: brain mechanisms of attention. Q J Exp Psychol. 59:2–27.

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

Eimer M, Shröger E. 1998. ERP effects of inter-modal attention and cross-modal links in spatial attention. Psychophysiology. 35:313–337.

Greene AJ, Easton RD, LaShell LS. 2001. Visual-auditory events: cross-modal perceptual priming and recognition memory. Conscious Cogn. 10:425–435.

Guthrie D, Buchwald JS. 1991. Significance testing of difference potentials. Psychophysiology. 28:240–244.

Hansen JC, Hillyard SA. 1980. Endogenous brain potentials associated with selective auditory attention. Electroencephalogr Clin Neurophysiol. 499:277–290.

Hillyard SA, Simpson GV, Woods DL, VanVoorhis S, Munte TF. 1984. Event-related brain potentials and selective attention to different modalities. In: Reinoso-Suarez F, Aimone-Marsan C, editors. Cortical integration. New York: Raven Press. p. 395–413.

James TW, Humphrey GK, Gati JS, Servos P, Menon RS, Goodale MA. 2002. Haptic study of three-dimensional objects activates extrastriate visual areas. Neuropsychologia. 40:1706–1714.

Kutas M, Hillyard SA. 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. Science. 207:203–205.

Martinez A, Teder-Sälejärvi WA, Hillyard SA. 2007. Spatial attention facilitates selection of illusory objects: evidence from event-related brain potentials. Brain Res. 1139:143–152.

Martinez A, Teder-Salejarvi W, Vazquez M, Molholm S, Foxe JJ, Javitt DC, Di Russo F, Worden MS, Hillyard SA. 2006. Objects are highlighted by spatial attention. J Cogn Neurosci. 18:298–310.

McCarthy G, Wood CC. 1985. Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. Electroencephalogr Clin Neurophysiol. 62:203–208.

McDonald JJ, Teder-Sälejärvi WA, Di Russo F, Hillyard SA. 2003. Neural substrates of perceptual enhancement by cross-modal spatial attention. J Cogn Neurosci. 15:10–19.

McDonald JJ, Teder-Sälejärvi WA, Hillyard SA. 2000. Involuntary orienting to sound improves visual perception. Nature. 407:906–908.

Melcher D, Vidnyanszky Z. 2006. Subthreshold features of objects: unseen but not unbound. Vision Res. 46:1863–1867.

Molholm S, Martinez A, Shpaner M, Foxe JJ. 2007. Object-based attention is multisensory: co-activation of an object's representations in ignored sensory modalities. Euro J Neurosci. 26:499–509.

Molholm S, Ritter W, Javitt DC, Foxe JJ. 2004. Multisensory visual–auditory object recognition in humans: a high-density electrical mapping study. Cereb Cortex. 14:452–465.

Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ. 2002. Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Cogn Brain Res. 14:115–128.

Murray MM, Molholm S, Michel CM, Heslenfeld DJ, Ritter W, Javitt DC, Schroeder CE, Foxe JJ. 2005. Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. Cereb Cortex. 15(7):963–974.

Näätänen R, Gaillard AWK, Mäntysalo S. 1978. Early selective-attention effect on evoked potential reinterpreted. Acta Psychol. 42:313–329.

O'Craven KM, Downing PE, Kanwisher N. 1999. fMRI evidence for objects as the units of attentional selection. Nature. 401:584–587.

Robertson LC. 2003. Binding, spatial attention and perceptual awareness. Nat Rev Neurosci. 4:93–102.

Saenz M, Buracas GT, Boynton GM. 2003. Global feature-based attention for motion and color. Vision Res. 43:629–637.

Schoenfeld MA, Tempelmann C, Martinez A, Hopf JM, Sattler C, Heinze HJ, Hillyard SA. 2003. Dynamics of feature binding during object-selective attention. Proc Natl Acad Sci USA. 100: 11806–11811.

Serences JT, Yantis S. 2006. Selective visual attention and perceptual coherence. Trends Cogn Sci. 10(1):38–45.

Stroop JR. 1935. Studies of interference in serial verbal reactions. J Exp Psychol. 18:643–662.

Talsma D, Doty TJ, Woldorff MG. 2007. Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? Cereb Cortex. 17:679–690.

Talsma D, Kok A. 2001. Nonspatial intermodal selective attention is mediated by sensory brain areas: evidence from event-related potentials. Psychophysiology. 38:736–751.

Teder-Sälejärvi WA, Munte TF, Franz-Jurgen S, Hillyard SA. 1999. Intramodal and cross-modal spatial attention to auditory and visual stimuli. An event-related brain potential study. Cogn Brain Res. 8:327–343.

Treisman A, Gelade G. 1980. A feature-integration theory of attention. Cogn Psychol. 12:97–136.

Treue S, Martinez-Trujillo JC. 1996. Feature-based attention influences motion processing gain in macaque visual cortex. Nature. 399:575–579.

Van der Burg E, Olivers CNL, Bronkhorst AW, Theeuwes J. 2008. Pip and pop: nonspatial auditory signals improve spatial visual search. J Exp Psychol Hum Percept Perform. 34(5):1053–1065.

Wylie GR, Javitt DC, Foxe JJ. 2003. Task switching: a high-density electrical mapping study. Neuroimage. 20:2322–2342.

Wylie GR, Javitt DC, Foxe JJ. 2004. Don't think of a white bear: an investigation of the effects of sequential instructional sets on cortical activity in a task-switching paradigm. Hum Brain Mapp. 21(4):279–297.

Wylie GR, Javitt DC, Foxe JJ. 2006. Jumping the gun: is effective preparation contingent upon anticipatory activation in task-relevant neural circuitry. Cereb Cortex. 16(3):394–404.