Journal of Experimental Psychology:
Human Perception and Performance
2013, Vol. 39, No. 3, 861–871

0096-1523/13/$12.00 DOI: 10.1037/a0030517

Contingent Attentional Capture by Conceptually Relevant Images

Brad Wyble
The Pennsylvania State University

Charles Folk
Villanova University

Mary C. Potter
Massachusetts Institute of Technology

Attentional capture is an unintentional shift of visuospatial attention to the location of a distractor that is either highly salient, or relevant to the current task set. The latter situation is referred to as contingent capture, in that the effect is contingent on a match between characteristics of the stimuli and the task-defined attentional-control settings of the viewer. Contingent capture has been demonstrated for low-level features, such as color, motion, and orientation. In the present paper we show that contingent capture can also occur for conceptual information at the superordinate level (e.g., sports equipment, marine animal, dessert food). This effect occurs rapidly (i.e., within 200 ms), is a spatial form of attention, and is contingent on attentional-control settings that change on each trial, suggesting that natural images can be decoded into their conceptual meaning to drive shifts of attention within the time course of a single fixation.

Keywords: contingent capture, attention, natural images, conceptual information

The information content of complex visual scenes typically exceeds the resources available to the human visual system, especially if the eyes are moving several times per second. One way of alleviating this fundamental limitation is to rapidly orient the available resources to task-relevant stimuli. Research suggests that this rapid orienting is accomplished by virtue of a "configurable" attention-allocation system (Folk, Remington, & Johnston, 1992, 1993; Folk & Remington, 1998). According to the contingent attentional-capture (CAC) theory, when asked to look for a target defined by a particular visual feature, the attention-allocation system is configured to prioritize processing of that feature, resulting in an involuntary shift of attentional resources to the location of any stimuli carrying that feature. This effect has been demonstrated for simple visual features, such as color, onset, and motion (Folk, Remington, & Wright, 1994; Folk & Remington, 1998) in addition to more abstract properties such as "singleton-ness" (i.e., when a stimulus differs from its neighbors; Folk & Anderson, 2010; Bacon & Egeth, 1994) and target-distractor relations (e.g., "redder;" Becker, Folk, & Remington, 2010). The data supporting the contingent-capture effect suggest that the focus of spatial attention is automatically reoriented toward a stimulus containing the feature that participants are looking for, even when doing so is detrimental to the current task.

In a classic demonstration of contingent capture (Folk, Leber, & Egeth, 2002), participants searched a central stream of colored letters for a letter of a particular color (e.g., red) and were required to report the identity of the letter at the end of the sequence. At variable times prior to the presentation of the target, four parafoveal hash marks appeared around the central stream. When one of the hash marks (referred to as the distractor) was the same color as the target, performance dropped significantly. No such effect was found when the distractor was not the same color as the target. Thus, distractors that carry task-relevant features have a detrimental impact on participants' ability to attend to targets in the central rapid serial visual presentation (RSVP) stream if they occur temporally in advance of the target. This is consistent with the basic premise of CAC, that stimuli matching current attentional-control settings produce involuntary shifts of attention.

Two important aspects of contingent capture are its rapid timing and its spatial specificity. In Folk et al. (2002), as well as numerous other studies (Folk, Leber, & Egeth, 2008; Folk, Ester, & Troemel, 2009), contingent-capture effects peak very rapidly (within about 150–200 ms of the onset of the capturing stimulus) and abate gradually over the course of 600–800 ms (Folk et al., 2008). Furthermore, the capture effect evokes a spatially specific enhancement of processing at the location of the capturing stimulus. For example, in Experiment 4 of Folk et al. (2002), a gray prime letter of the same identity as the central target was presented immediately after the distractor display, and appeared at the same or different location as the target-colored distractor. Priming effects were only observed when the prime appeared at the same location as the distractor, demonstrating that the contingent-capture effect is spatially specific.

This article was published Online First November 19, 2012.

Brad Wyble, Department of Psychology, Center for Brain, Behavior and Cognition, The Pennsylvania State University; Charles Folk, Department of Psychology, Villanova University; Mary C. Potter, Brain and Cognitive Sciences Department, Massachusetts Institute of Technology.

This work was supported by Grant No. MH47432 from the National Institute of Mental Health. We thank Aminah Assabahi, Amilcar Bautista, Matthew Bennett, Jr, Laura Braun, Colin Flowers, Jenessa Holder, Nick Jones, and Winston McCarty for research assistance.

Correspondence concerning this article should be addressed to Brad Wyble, The Pennsylvania State University, 140 Moore Building, University Park, PA 16802. E-mail: bwyble@gmail.com

A good deal of research has focused on the nature and flexibility of the attentional-control settings associated with CAC (e.g., Bacon & Egeth, 1992; Folk & Anderson, 2010; Folk et al., 1994; Folk & Remington, 1998). In all of these cases, it has been assumed that attentional-control settings are associated with simple visual features (e.g., color, size, shape, brightness, orientation, etc.) that are processed and represented in preattentive "feature maps." These feature maps provide retinotopic representations of basic features present in the visual array (Cave & Wolfe, 1990; Treisman & Gelade, 1980). For example, an attentional-control setting for "red" may be instantiated by increasing the "gain" associated with signals from the "red" color map, which would then increase the probability that any stimuli carrying that color would elicit a shift in attentional resources.

However, it is logically possible that any type of information that is processed preattentively (not just the simple features known to populate feature maps) can serve as a potential candidate for an attentional-control setting. One such candidate, to which we will now turn, is conceptual or semantic information associated with natural visual images.

Conceptual Processing

The visual system is able to extract conceptual information from an image at great speed (Potter, 1976). Even at stimulus-onset asynchronies (SOAs) as short as 113 ms per item, participants can detect target images specified by a conceptual description of its gist, such as "a road with cars" or "a girl sitting in bed." This rapid conceptual understanding of an image—the kinds of objects and the type of place or scene—is accomplished by tapping into the visual system's past experience with visual input.

Viewers can also report specific targets defined by a superordinate category, such as animal or furniture, when shown pictured exemplars such as a moose or a sofa (Potter, Wyble, Pandav, & Olejarczyk, 2010). This study used RSVP of natural images in an attentional-blink paradigm and found that the temporal dynamics of attention, as triggered by natural scenes, are similar to the temporal dynamics found with simpler stimuli, such as letters and digits. In this paradigm, participants saw one of 29 target specifications, such as dinner food or marine animal, after which an RSVP sequence of natural images at 107 ms was presented containing two novel targets (T1 and T2) separated by one, two or four lags. The task was to report the specific identities of each of the targets. The proportion of correct reports of T2 followed the traditional attentional pattern, with Lag-1 sparing (i.e., minimal reduction in T2 reports at Lag 1), an attentional blink at Lag 2 (i.e., substantial reduction in T2 reports), and subsequent recovery with increasing lag.

According to many accounts of the attentional blink (Chun & Potter, 1995; Olivers & Meeter, 2008; Bowman & Wyble, 2007; Wyble, Bowman, & Nieuwenstein, 2009), Lag-1 sparing is the result of a transient attentional process (Nakayama & Mackeben, 1989), which is triggered in response to a detected target and peaks within about 150 ms after the onset of the target. This mechanism explains the excellent report of T2 when it occurs in the Lag-1 position. It has also been demonstrated that the categorical target set used in previous attentional-blink studies (i.e., letter vs. digit) is capable of triggering rapid shifts of attention on the same time scale as the contingent-capture effect (Wyble, Bowman, & Potter, 2009).

These studies provide evidence for the rapid, voluntary allocation of attention to images or stimuli containing task-relevant conceptual and categorical information. There is also evidence that images containing certain classes of conceptual information can elicit the involuntary allocation of attention (i.e., attentional capture). Specifically, there is a growing literature suggesting that images with threatening content or negative emotional valence can capture attention (e.g., Fox et al., 2000; Fox, Russon, Bowles, & Dutton, 2001; Most, Chun, Widders, & Zald, 2005; Nummenmaa, Hyönä, & Calvo, 2009; Öhman, Flykt, & Estevez, 2001; Rinck, Becker, Kellerman, & Roth, 2003; Schmidt, Belopolsky, & Theeuwes, 2012). For example, using a visual-search task, Öhman et al. (2001) found that fear-relevant targets (e.g., spiders, snakes) among neutral distractors (e.g., flowers) are detected more rapidly than neutral targets. More relevant to the current work, Most et al. (2005) found that discriminating the orientation of a target scene in an RSVP stream is impaired when the target is preceded (at Lag 2) by a task irrelevant image with highly negative emotional content (e.g., graphic violence, mutilation, etc.). Such results have been interpreted as reflecting the activity of a specialized emotional system that has evolved to automatically orient the observer to environmental stimuli based on threat-related content rather physical salience.

In the present work, we explore whether the capture of attention by conceptual information may extend beyond threat-related images. Specifically, we test the hypothesis that nonthreat-related images can capture attention by virtue of conceptual content, provided that the content is relevant to task goals. Is it possible to set the attention-allocation system for a conceptual category in the same way it is possible to set for simple features like color? If so, can capture by conceptually relevant stimuli occur in RSVP streams of the same speed as capture by simpler features such as color? And do such effects have a spatial component, as in Folk et al. (2002)? To test these questions, three experiments were run, which were similar to the design of experiments in Folk et al.(2002) using the set of natural images in Potter et al. (2010).

Experiment 1

In a design modeled after Experiment 1 of Folk et al. (2002), participants monitored a central RSVP stream of images for a target image. A written target-category name (e.g., amusement-park ride) was presented at the beginning of each sequence. The task was to report the identity of the one image in the sequence (i.e., the target) that was an exemplar of the target category by typing the name of the image (e.g., bumper cars). Two frames prior to the target, distractor images appeared in the periphery above and below the central stream. On critical trials, one of the distractor images was a different exemplar from the target category (e.g., a ferris wheel). If it is possible to instantiate an attentional-control setting for a conceptual category, then according to CAC theory, any exemplar from that category should capture attention. Thus, the presence of a parafoveal image that belongs to the target category should capture attention and result in a decrement in target report.

Method

Participants. Twelve studentsfrom the Villanova University research participation pool participated in the experiment in exchange for credit for a course requirement. All participants had normal or corrected-to-normal vision and none reported being color-blind. One subject was replaced because of accuracy more than two standard deviations below that of the other participants.

Apparatus. The experiment was run using Matlab 7 with Psychophysics Toolbox extensions (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007) on a Dell Optiplex 780 PC computer. Stimuli were presented on a Sony Trinitron Multiscan 500PS CRT monitor, with the screen set to 1024 × 768 resolution.

Stimuli. The stimuli, from a set used by Potter et al. (2010), consisted of colored photographs of single objects in their natural settings and pictured scenes containing multiple objects and contextual settings. The pictures were natural images downloaded from Google Images. There were a total of 160 target pictures and 561 distractor pictures. The pictures were modified in Adobe Photoshop CS and resized to 300 × 200 pixels. All writing was removed or "air-brushed" from the pictures, as were other unwanted visual features. Pairs of target images were selected from 29 superordinate categories of objects such as fruit, vehicle, body part, or cleaning product; the pictured exemplars were typical of basic-level examples of the superordinate categories, for example, banana, boat, ear, or broom. See the Appendix for a list of the names of all stimulus categories and examples.[1] For each of the 80 trials, two unique target images from one superordinate category were chosen, avoiding closely related or similar-looking exemplars of the category. For each participant and for each trial, one of the two target-category pictures was randomly selected as the central RSVP target, and the other was used as the same-category-distractor on that trial.

Distractor images could be of scenes containing people, animals, natural landscapes, and images of single or multiple objects. The distractor images within each trial were inspected to ensure that none of them were a partial match to the target category. Of the 561 distractors, 242 of them were presented once per subject and 319 distractors were presented twice, although never in the same trial.

Procedure. Each trial contained a conceptual-target specification (e.g., body part), two target images matching that specification, and 10 distractor images. Each trial began with a specification of the target category for 600 ms on a gray screen, followed immediately by the presentation of a sequence of nine images at 129 ms per image, with no inter-stimulus interval (ISI). Each trial had one target image, which was presented randomly in Position 5, 6, 7, or 8. Two frames prior to the target image (i.e., Lag 2), two additional distractor images were presented above and below the central RSVP (see Figure 1). The central images measured 3.44 × 2.29° visual angle at a distance of 50 cm. The parafoveal images were 4.58 × 2.98° visual angle and centered 3.44° above and below the central RSVP stream (measured from the center of the central images to the center of the parafoveal images). On half of the trials, randomly selected for each subject, one of these two images was a same-category distractor, and appeared randomly above or below the RSVP stream. On the other half of the trials, neither distractor was related to the target category.

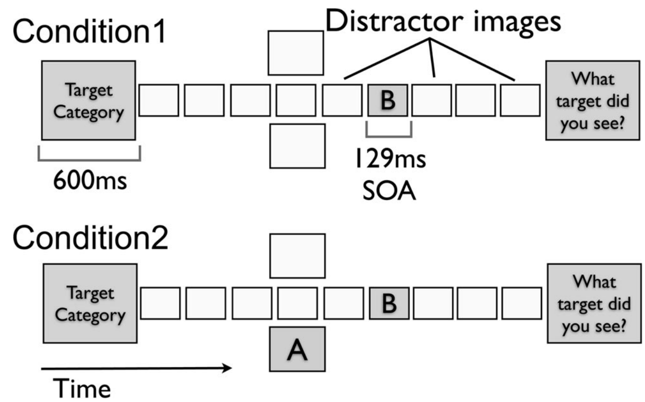

Figure 1. Design of Experiment 1. Each trial began with a screen showing the target category, after which nine images were shown in RSVP. A and B denote the two images that match the target category. The other rectangles denote images that do not match the target category. Two larger rectangles denote the distractor images shown above and below the RSVP.

Immediately after the stream was over, participants were shown the prompt "What target did you see?" and were allowed to type a response which could include multiple words. Backspacing to correct the response was permitted. No feedback was given.

Each participant saw two blocks of the same 80 trials. Each trial consisted of a specific sequence of distractors, with the exception that the two images that matched the target category could be placed in one of the parafoveal positions, or the central RSVP stream. Which of the two images was used as a target and which was a distractor was randomly chosen for each trial, separately for each participant. The order of trials in each block was randomly shuffled, such that each subject saw a different ordering. All parameters, including trial order, and which of the two same-category images was the target were rerandomized for the second block.

Scoring. Participants' responses were scored by volunteers who were blind to the condition of each trial, except that they knew the name of the target category. Scorers were instructed to score a mismatching response as correct if it was the same category, very closely related such that it could be easily mistaken (e.g., papaya and mango), a spelling error but possibly correct, or a close/correct description. A response was scored incorrect if it was too general, nonspecific, vague (e.g., bird, which was the category name), or same category but unrelated (e.g., elephant instead of tiger).

Results

The results of this study provide clear evidence of attentional capture by same-category distractors (see Figure 2). For trials in which the same-category distractor was not presented, participants identified the target on .77 (SE = .017) of the trials. On trials containing a same-category distractor, this probability was reduced to .58 (SE = .031). This difference was significant, $F(1, 11) = 52.786$, $p < .0001$, $\eta_p^2 = .83$. This result shows that targets were reported less often when one of the two distractors outside of the

[1] For examples of stimuli, please visit http://www.bradwyble.com/research/materials/ConceptualCapture

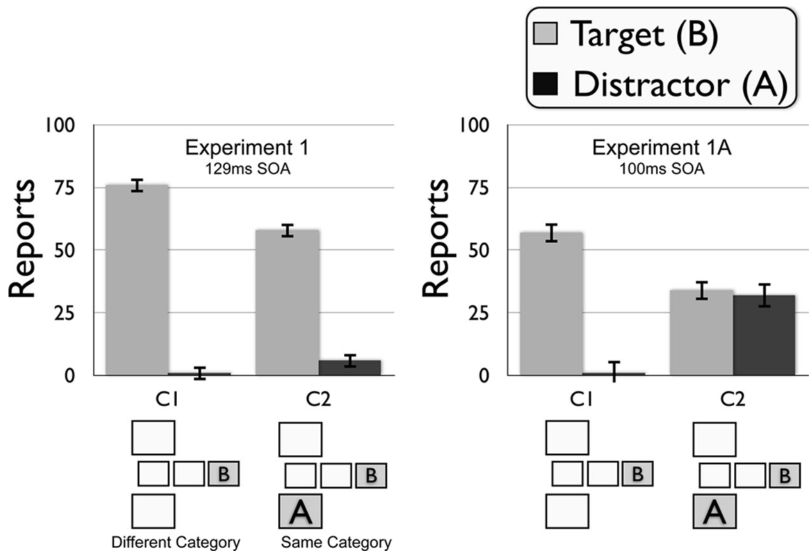

Figure 2. Results from Experiment 1 (left panel) and Experiment 1a (right panel). The vertical axis denotes the percentage of trials in which participants reported the name of the target (Image B) and the same-category distractor (Image A) in each of the two conditions. Bars indicate standard error. The diagrams below the bar illustrate the conditions.

RSVP stream was of the same category as the target, and this finding supports the hypothesis that natural images can capture attention by virtue of their conceptual similarity to the target. Participants would sometimes report the same-category distractor in lieu of the target. Such errors happened on .09 (SE = .034) of the same-category trials.

Because each individual saw two copies of the 80 trials, we could compare performance on the second presentation of the same trials to see if they were notably different from the first presentation, thereby assessing whether familiarity with the targets enhances or reduces capture. In the first block, participants identified the targets on .77 (SE = .026) of the different category trials and .53 (SE = .033) of the same-category trials. In the second block of the same trials, participants reported the target on .77 (SE = .015) and .63 (SE = .04) respectively for the same conditions. These results indicate that the capture effect was present on the first exposure of each of the target stimuli and suggest that participants may have gotten better at ignoring the distractors on the second block. A 2 factor ANOVA found significant effects of condition, $F(1, 11) = 52.8$, $p < .0001$, $\eta_p^2 = .83$, block, $F(1, 11) = 6.3$, $p < .03$, $\eta_p^2 = .36$, and a nonsignificant interaction, F (1,11) = 3.3, $p < .1$, $\eta_p^2 = .23$.

Some of the unrelated distractors were repeated across trials within each block, but the target and same-category distractor images were never repeated during a block. Therefore, it is possible that when the same-category distractor was presented simultaneously with an image that was more familiar to the subject, attention was drawn to the categorical stimulus by its relative novelty, rather than its conceptual content. To test whether this explanation could account for the results, a follow-up analysis included only trials in which the parafoveal distractors were presented once per experiment block. This restriction equated the novelty of the same-category distractor with the accompanying distractor, and thus eliminated the possible confound of novelty as a driver of capture. This analysis yielded target reports that were highly similar to the proportion of reports from the preceding analysis (.74 vs. .57 in this analysis, and .77 vs. .58 in the original analysis). Thus, the relative novelty of the accompanying distractor did not contribute to the capture effect.

Experiment 1A

Experiment 1 yielded results that were consistent with the CAC theory, with a conceptual category serving as the attentional set. When a target image in the central stream was preceded by a same-category distractor in the periphery, report of that target image was significantly less accurate. In that experiment, the SOA of the stream was 129 ms, which is significantly slower than conventional RSVP speeds used in attentional-blink and conceptual-capture experiments. To achieve the combined aims of replicating Experiment 1 and also to determine whether natural images elicit a contingent-capture effect at more conventional RSVP rates, Experiment 1A duplicated the same design except with a 100 ms SOA.

Method

All methods were the same as in Experiment 1 except that in 1A, there were 14 participants from Villanova University, and the duration of the stimuli was shortened to 100 ms SOA.

Results

The results of Experiment 1A replicated those of Experiment 1 (see Figure 2), with higher accuracy on trials without same-category distractors ($M = .57$, $SE = .04$) than on trials with them ($M = .34$, $SE = .044$), which was significant ($F_{1,\ 13} = 45.7$, $p < .0001$, $\eta_p^2 = .78$). One significant difference from the results of

Experiment 1 is that erroneous report of same-category distractors was considerably higher at 100 ms, suggesting that participants had a more difficult time registering the location of the task-relevant target in the middle of the screen. In this experiment, participants erroneously reported the same-category distractor on .32 of the trials (SE = .055), compared with .09 (SE = .034) in Experiment 1.

Discussion

These results demonstrate that the presence of a distractor containing task-relevant conceptual content decreases accuracy of reporting a target. The underlying theoretical interpretation of these results is that attention was shifted toward the location of the same-category distractor, and furthermore, that the visual system produced this shift contingently as a result of computing a match between the conceptual meaning of the same-category distractor and the target set. The following two experiments will address these points.

Experiment 2

In the first two experiments, both the target and distractors set had substantial heterogeneity in terms of colors, textures, luminance values, and other low-level stimulus characteristics. It was nevertheless possible that there were systematic low-level differences between the target and distractor images because the target images were hand-picked to match the target specifications, whereas the distractors were chosen because they did not match these target-category specifications. Despite the best of intentions in eliminating low-level differences between these picture sets, the difference in selection strategy could have inadvertently created a target image set that is more salient than the distractor set for reasons unrelated to task relevance. If this were the case, then the capture effects we observed could have been driven entirely by bottom-up salience, rather than being contingent on a match between conceptual content and the target set for a particular trial. To investigate this alternative hypothesis, we ran an additional experiment in which stimuli that were part of the target set on a given trial were reassigned into a trial with a different target set to see if they could still capture attention.

As an example, on a given trial, a superordinate target category might be amusement ride, with the target images being bumper cars and a ferris wheel. On another trial the superordinate target category might be marine animal, with targets being a dolphin and a sea horse. To create a *reassigned* trial, the sea horse would be used as a critical distractor in a trial with the amusement ride target set. So if the sea horse captures attention only in the marine animal trials, then the match with the target category would be found to be essential for eliciting the effect. However if the sea horse is somehow more salient than the other distractors (e.g., it has a pop-out color) then it will capture attention even when the target category is amusement ride.

Method

Experiment 2's method was similar to that of Experiment 1a except as follows: 21 participants from the Villanova University research participation pool took part in this experiment. Two were excluded for failing to understand the instructions, leaving a total of 19 participants.

Design and Procedure

Prior to data collection, for each trial, the two images that matched the target set were manually assigned to a different trial to serve as distractors on a reassigned trial. This was done manually using a spreadsheet in which the images were accessed only by index number. At the beginning of each subject's experiment, half of the 80 trials were randomly selected to be reassigned trials (i.e., they would use these reassigned images in place of the same-category distractor). Thus, on these reassigned trials, images that would have been same-category distractors in their original context were used as distractors for trials in which they did not match the target set. As a further control, the reassignment manipulation also copied the parafoveal distractor, which had originally accompanied the same-category distractor. This was done to preserve the original pairing of the two parafoveal images.

Thus, there were two conditions in this experiment. In the intact condition, one of the two parafoveal distractors matched the target set for the current trial. In the reassigned condition, one of the two parafoveal distractors matched the target set for a different trial.

Results

The results of Experiment 2 (see Figure 3) were consistent with the idea that the capture effect observed in Experiments 1 and 1a was the result of a match between the target set of each trial, and the conceptual information in the same-category distractor. In this experiment, accuracy of detecting the central RSVP target was reduced in the intact trials (M = .43, SE = .03) relative to the trials

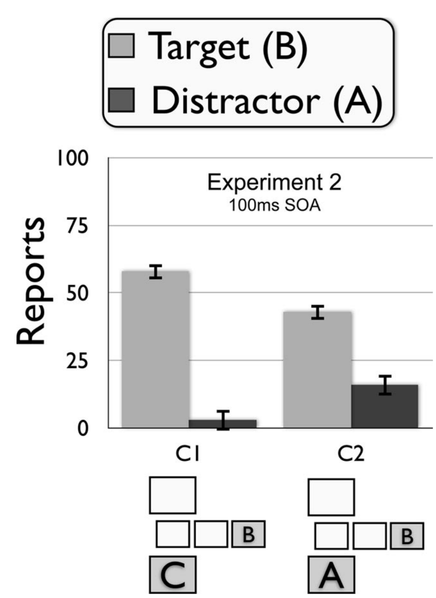

Figure 3. Results from Experiment 2. In this diagram, A refers to a same-category distractor from an intact pair (i.e. it matched the target specification of the current trial). C refers to an image that was used as a same-category distractor in Experiments 1, 1a, and 2, but in this experiment, had been reassigned to a trial with a different conceptual target set.

in which the category distractors were reassigned (M = .58, SE = .04) and this effect was highly significant ($F_{1, 18}$ = 40.5, p < .0001, η_p^2 = .69). In comparing these results directly to those of Experiment 1a, which had the same SOA, performance in the mismatch trials was nearly identical to the different-category distractor condition in that experiment (M = .57 in Experiment 3 vs. M = .58 in Experiment 1a). These results clearly suggest that the attentional settings of the subject play a role in producing the capture effects observed in the preceding experiments.

Discussion

Although the results of the preceding experiments provide evidence for the capture of spatial attention by task-relevant conceptual information, there is another way of accounting for the results without invoking a spatially specific form of attentional capture. Specifically, it is possible that participants have a difficult time keeping their focus tightly constrained onto the central RSVP. Under this hypothesis (see Figure 4), participants encode both the target and the same-category distractor, and have difficulty binding each one to its location. Thus, participants become confused about which image was presented in the central location and have to guess at which image was the target on some trials. Experiment 3 was designed to test this possible explanation by providing more direct evidence that a spatial-capture effect is elicited by the same-category distractor.

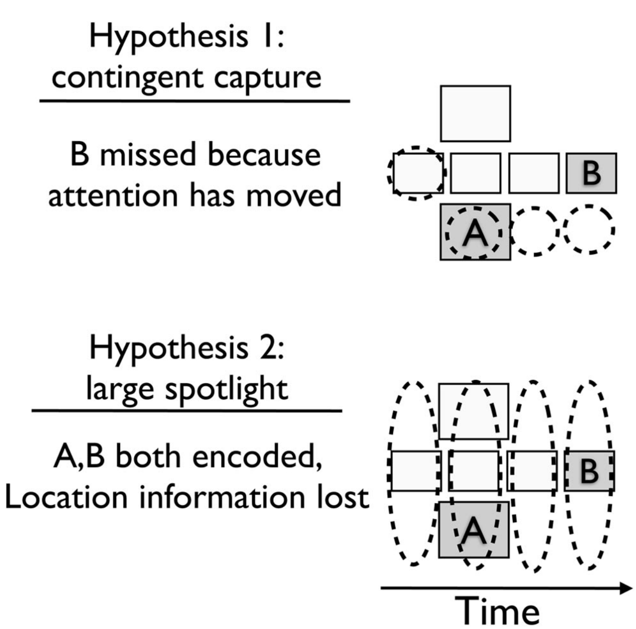

Figure 4. An illustration of two competing hypotheses for the observed data. The contingent-capture hypothesis suggests that attention is spatially shifted by the same-category distractor (A). An alternative hypothesis, addressed in Experiment 3, proposes that the attentional focus includes the parafoveal images despite our instructions to attend to the center stream. If participants failed to encode the location of the images, they might have been confused about which image to report, and this would have produced an apparent-capture effect.

Experiment 3

Folk et al. (2002) used a priming experiment (Experiment 4 in their paper) to test whether attention was being captured to the location of the same-color distractor. They found positive priming of the target when a copy of the upcoming target immediately followed the same-color distractor in the same location. That is, giving a preview of the target immediately after the distractor mitigated the capture effect. Experiment 3 implemented a similar paradigm using a sequence of images. In the critical condition, a copy of the central target image (i.e., the prime) was presented immediately after the distractor images, and could appear at the location of the same-category distractor, or the different-category distractor. If same-category distractors attract spatial attention to their location, then the target prime should enhance report of the central target to a greater extent when it appears at the location of the same-category distractor relative to the location of the different-category distractor.

Method

Experiment 3 was similar to Experiment 1a except as follows: 41 students, aged 18–24 from the Syracuse University research participation pool took part in this experiment. Participants received course credit in exchange for participation.

The RSVP stream was composed of eight stimuli presented at a rate of 93 ms per image. The target image was presented in Position 5, 6 or 7. The images presented outside of the RSVP stream now consisted of two distractors presented two positions prior to the target, and two additional distractors presented one position prior to the target.

There were four conditions, as shown in Figure 5. In Condition 1, none of the four parafoveal distractors was of the target category. In Condition 2, one of the distractors in the first set of two distractors was a same-category distractor. In Conditions 3 and 4, both images belonging to the target set were presented as distractors. The first distractor was always the nontarget image (A in Figure 5) and the second distractor was the same as the target image 0 (B in Figure 5). In Condition 3, the same-category distractor and the target prime were in the same location and in Condition 4 they were in different locations. There were 80 trials, with 20 trials in each condition, intermixed randomly for each participant.

Results and Discussion

The results are illustrated in Figure 5. An omnibus ANOVA across all four conditions found a significant effect of condition ($F_{3, 40}$ = 52.16, p < .0001, η_p^2 = .56). In a planned analysis of Conditions 1 and 2, the results replicated the attentional-capture effect from Experiment 1a, with a mean of .58 (SE = .024) in Condition 1 and .43 (SE = .023) in Condition 2. This difference was significant ($F_{1, 40}$ = 61.484, p < .0001, η_p^2 = .61). The critical new comparison is between Conditions 3 and 4 in which the target-prime image is shown in the same spatial position (Condition 3) or a different position (Condition 4) from the same-category distractor. Of note, Condition-3 accuracy (M = .67, SE = .02) was higher than Condition-4 accuracy (M = .59, SE = .02, $F_{1, 40}$ = 21.61, p < .0001, η_p^2 = .35), showing that there is a spatial component to this attentional capture effect.

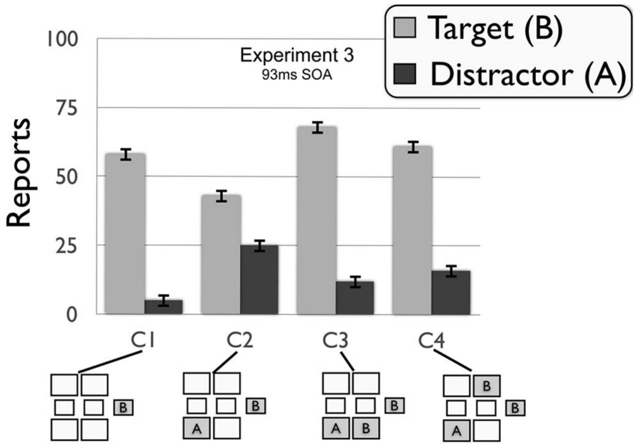

Figure 5. Results from the four conditions of Experiment 3. Conditions 1 and 2 replicate the conceptual-capture effect. Conditions 3 and 4 demonstrate that there is a spatial component, with enhanced report of B when it directly follows the same-category distractor (A), rather than appearing in a different location.

Another important aspect of the results of Experiment 3 is that Conditions 1 and 2 demonstrate conceptual capture even by a masked distractor. In Experiments 1, 1A, and 2, the same-category distractor was unmasked, and thus might have produced a strong iconic representation. However, in Condition 2 of Experiment 3, the first pair of parafoveal distractors is always masked by the following pair of distractors, and yet the capture effect is still clearly observed.

General Discussion

The results of the present experiments show that the ability to report a conceptually specified target in a central RSVP stream is impaired when preceded by a parafoveal distractor matching the conceptual description of the target. This represents a contingent-capture effect, in which the attentional-control settings of the participants were configured to a conceptual category such as amusement ride or dinner food. The effect did not depend on familiarity with specific visual features, since it could be observed even during the first block in Experiment 1, when each target image was novel to the participant. Experiment 2 provided evidence that the conceptual content of the same-category distractors (rather than low-level features) is responsible for the effect. Finally, Experiment 3 demonstrated that that there is a spatial component to this conceptually driven capture effect by presenting a target prime in the same or a different location as the capturing distractor. Performance in identifying the target was better when the prime followed in the same location as the same-category distractor than when it appeared in a different location, which demonstrates that attention was captured at the location of the same-category distractor.

Contingent-capture effects have been well established for attentional sets defined by single visual features such as color and motion cues (Folk et al., 1994; Folk & Remington, 1998; Becker et al., 2010). The present results expand our understanding of attentional-control settings to include relatively abstract conceptual information for a large variety of categories, specified anew on each trial.

These results have implications for our understanding of the flexibility of attentional-control settings. What they suggest is that attentional set can be established at the level of concepts that are abstracted from specific visual features. Furthermore, this conceptual attentional set can be used to trigger the deployment of attention at a rapid time scale, as evidenced in Experiment 3 in the comparison of Conditions 3 and 4. In this experiment, a same-category distractor affected the effectiveness of the following target prime at an SOA of 93 ms, an SOA at which attentional facilitation has been found to occur for simpler visual stimuli such as letters and digits (Wyble, Bowman, & Potter, 2009).

The present results also illustrate the malleability of attentional-control settings. In Experiment 2, performance in the reassigned condition was at nearly the same level as in the different-distractor condition of Experiment 1a. If attentional settings carried over substantially from one trial to the next, we would expect the reassigned images to have substantial capture effects due to carryover of attentional set across trials.

Relation to Other Conceptual-Capture Work

In the present results, attention is drawn away from the central RSVP by a distractor that is known to be task-irrelevant because of its location, but is nevertheless a source of distraction because it is a member of the target set. A related finding that agrees with our results is a set of studies by Sulman and Sanocki (2011), who found capture effects by scenes that matched the conceptual context of the target set in an RSVP experiment. For example, while an individual is looking for a tennis ball, his or her ability to detect it was reduced if a picture of a tennis court appeared ahead of it in the RSVP stream.

Other recent work has looked for a similar effect in a dot-probe task. In Vogt, De Houwer, Moors, Van Damme, and Crombez (2010), participants were asked to adopt a goal set for producing speeded responses to a particular word. While this task set was engaged, participants were given a dot-probe task with either the task-relevant word, or a synonym of it, used as a cue. The task-relevant word itself, but not the synonym, was able to elicit a disengagement deficit at an SOA of 250 ms. In a follow-up paper (Vogt, De Houwer, & Moors, 2011), the response task was to respond to either a word or a picture, and the results were similar. These results illustrate a case of a highly specific attentional set, and find that there is no carryover to other stimuli in the semantic neighborhood of that specific attentional set. In contrast, the attention sets in the present paper are superordinate-level categories, and the capturing stimuli fit within that category.

Capture by Features or Capture by Concept?

Evans and Treisman (2005) have suggested that natural scenes containing a particular kind of target can be detected based on their individual features, and they describe the case of discriminating animals from humans in RSVP sequences using individual features that are characteristic of animals, such as wings, beaks, fur, or gills. In a similar vein, Levin, Takarae, Miner, and Keil (2001) describe the potential contribution of rectilinear feature detection to discriminate animals from human-made artifacts in visual search of line drawings.

Therefore, in the present experiments, one might suppose that the visual system would rapidly detect the relevance of stimuli to a conceptual category on the basis of particular features without requiring the visual system to compute the conceptual content of a stimulus.

There are, however, several aspects of the present experiments that make feature detection an unlikely explanation for this capture effect. First, the set of distractors was extremely heterogeneous. Thus, unlike the experiments in which participants could develop a clear distinction between animal or human (Evans & Treisman, 2005), or animal versus artifice (Levin et al., 2001), it is difficult to imagine a set of features that would discriminate targets from distractors even within a single trial. For example, in a trial for which the target category was *carrying-item*, the target images were a rectangular black briefcase and a round beige wicker basket. This trial had an average difference of .31 between the probability of report in conditions 1 and 2 of Experiment 3. Participants had never seen the target images before and thus could not have been expecting particular colors, orientations, textures, or even the specific kind of carrying item that would have been the target in the RSVP stream.

Another aspect of the present experiments that makes the feature-detection hypothesis an unlikely explanation is that the target category changed markedly from one trial to the next. There were 29 categories in all (see Appendix for a complete list), and participants in the experiment had little opportunity to fine tune their attentional set to a particular target specification over the course of multiple trials. However, it is possible that the overall effect is driven by a small number of categories that were discriminable from distractors based on low-level features, with the remainder of the trials having a null effect. To test this possibility we ran a secondary analysis on the results of Experiment 3, which had the most participants, to assess how the capture effect was distributed across the different categories. This analysis revealed that 26 of the 29 categories exhibited a numerically positive capture effect. Thus the capture effect is broadly distributed across the entire set of categories, rather than being confined to a small set of trials. This result suggests that the finding is a general phenomenon of conceptual set.

Neural Plausibility

Results from monkey neurophysiology support the possibility that conceptual information can be rapidly detected by the visual system. Hung, Kreiman, Poggio, and DiCarlo (2005) have demonstrated that within 125 ms of the onset of a visual stimulus, neurons within inferotemporal (IT) cortex of the primate visual system are responding selectively to the identity of that stimulus. This rapid identification process is thought to arise as a result of a sweep of feed-forward processing (Thorpe, Fize, & Marlot, 1996; Thorpe & Imbert, 1989) through the ventral visual pathway, in which a succession of cortical areas briefly process information and then pass it on to the next layer, from V1 to V2 and so on up to IT cortex Given these results, it is not inconceivable that capture effects can be elicited by natural images in as short a time as 200 ms.

Feature Binding and Preattentive Processing

The finding that conceptually relevant distractors trigger the deployment of attention implies that conceptual information was extracted from them in the absence of focal attention. Furthermore, at the time slot during which the capture occurred, the visual system was presented with three simultaneous images for 93 ms, which were then masked in Experiment 3. That capture occurred suggests parallel processing of multiple images. Thus, these results are consistent with the idea that processing of images can proceed in the absence of attention, at least for stimuli from familiar categories (see also Li, VanRullen, Koch, & Perona, 2002). Other work with rapidly presented images has also suggested that multiple images can be processed for conceptual information in parallel (Rousselet, Fabre-Thorpe, & Thorpe, 2002; Potter & Fox, 2009). Vanrullen (2009) has suggested that the features of highly familiar natural stimuli (e.g., images of animals) do not need to be bound in the same way as so-called "on-demand" feature binding (e.g., arbitrary shape–color combinations). Hommel and Colzato (2009) also argue that there is an important distinction between ad hoc binding of arbitrary feature conjunctions, and the binding of features comprised of highly familiar stimuli.

When viewed in this context, the present findings are not incompatible with conventional theories that feature binding requires attention, because such accounts are usually based on experiments that require binding of arbitrary colors and shapes across trials. We argue that, in the experiments presented here, participants could process the conceptual information within the images rapidly and in parallel, because they conformed to common visual types with which our participants had years of preexperimental experience.

Conclusion

In conclusion, the results of the present experiments clearly show that a verbal description of a conceptual category can configure attentional-control settings to be rapidly responsive to images that match a particular concept, even at the cost of being captured inadvertently by a stimulus in the wrong location. The capture effect would, however, be adaptive in a wide range of search tasks in which the searcher does not know the location of the missing object, such as a wallet or car keys.

References

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics, 55,* 485–496. doi:10.3758/BF03205306

Becker, S. I., Folk, C. L., & Remington, R. W. (2010). The role of relational information in contingent capture. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1460–1476. doi:10.1037/a0020370

Bowman, H., & Wyble, B. (2007). The simultaneous type, serial token model of temporal attention and working memory. *Psychological Review, 114,* 38–70. doi:10.1037/0033-295X.114.1.38

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

Cave, K. R., & Wolfe, J. M., (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology*. *22*. 225–271. doi:10.1016/0010-0285(90)90017-X

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 109–127. doi:10.1037/0096-1523.21.1.109

Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance 31* 1476–1492. doi:10.1037/0096-1523.31.6.1476

Folk, C. L., & Anderson, B. A. (2010). Target uncertainty effects in attentional capture: Singleton detection mode or multiple attentional-control settings? *Psychonomic Bulletin & Review, 17,* 421–426. doi:10.3758/PBR.17.3.421

Folk, C. L., Ester, E., & Troemel, K. (2009). How to keep attention from straying: Get engaged! *Psychonomic Bulletin & Review, 16,* 127–132. doi:10.3758/PBR.16.1.127

Folk, C. L., Leber, A., & Egeth, H. (2002). Made you blink! Contingent attentional capture produces a spatial blink. *Perception & Psychophysics, 64,* 741–753. doi:10.3758/BF03194741

Folk, C. L., Leber, A., & Egeth, H. (2008). Top-down control settings and the attentional blink: Evidence for nonspatial contingent capture. *Visual Cognition, 16,* 616–642. doi:10.1080/13506280601134018

Folk, C. L., & Remington, R. W. (1998). Selectivity in attentional capture by featural singletons: Evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 847–858. doi:10.1037/0096-1523.24.3.847

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional-control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1030–1044. doi:10.1037/0096-1523.18.4.1030

Folk, C. L., Remington, R. W., & Johnston, J. C. (1993). Contingent attentional capture: A reply to Yantis (1993). *Journal of Experimental Psychology: Human Perception and Performance, 19,* 682–685. doi:10.1037/0096-1523.19.3.682

Folk, C. L., Remington, R. W., & Wright, J. H. (1994). The structure of attentional control: Contingent attentional capture by apparent motion, abrupt onset, and color. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 317–329. doi:10.1037/0096-1523.20.2.317

Fox, E., Lester, V., Russo, R., Bowles, R. J., Pitchler, A., & Dutton, K. (2000). Facial expressions of emotion: Are angry faces detected more efficiently? *Cognition and Emotion, 14,* 61–92. doi:10.1080/026999300378996

Fox, E., Russo, R., Bowles, R., & Dutton, K. (2001). Do threatening stimuli draw or hold visual attention in subclinical anxiety. *Journal of Experimental Psychology: General, 130,* 681–700. doi:10.1037/0096-3445.130.4.681

Hommel, B., & Colzato, L. S. (2009). When an object is more than a binding of its features: Evidence for two mechanisms of visual feature integration. *Visual Cognition, 17,* 120–140. doi:10.1080/13506280802349787

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science, 310,* 863–866. doi:10.1126/science.1117593

Kleiner M., Brainard D., & Pelli D. (2007) What's new in Psychtoolbox-3? *Perception, 36* [ECVP Abstract Supplement]. Retrieved from http://www.perceptionweb.com/abstract.cgi?id=v070821

Levin, D. T., Takarae, Y., Miner, A., & Keil, F. C. (2001). Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. *Perception & Psychophysics, 63,* 676–697. doi:10.3758/BF03194429

Li, F.-F., VanRullen, R., Koch, C., & Perona, P. (2002). Natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 9596–9601. doi:10.1073/pnas.092277599

Most, S. B., Chun, M. M., Widders, D. M., & Zald, D. H. (2005). Attentional rubbernecking: Cognitive control and personality in emotion-induced blindness. *Psychonomic Bulletin & Review, 12,* 654–661. doi:10.3758/BF03196754

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research, 29,* 1631–1647. doi:10.1016/0042-6989(89)90144-2

Nummenmaa, L., Hyönä, J., & Calvo, M. G. (2009). Emotional scene content drives the saccade generation system reflexively. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 305–323. doi:10.1037/a0013626

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General, 130,* 466–478. doi:10.1037/0096-3445.130.3.466

Olivers, C. N., & Meeter, M. (2008). A booster/bouncer theory of visual attention. *Psychological Review*; *115,* 836–863. doi:10.1037/a0013395

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 509–522. doi:10.1037/0278-7393.2.5.509

Potter, M. C., & Fox, L. F. (2009). Detecting and remembering simultaneous pictures in RSVP. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 28–38. doi:10.1037/a0013624

Potter, M. C., Wyble, B., Pandav, R., & Olejarczyk, J. (2010). Picture detection in RSVP: Features or identity? *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1486–94. doi:10.1037/a0018730

Rinck, M., Becker, E. E., Kellerman, J., & Roth, W. T. (2003). Selective attention in anxiety: Distraction and enhancement in visual search. *Depression and Anxiety, 18,* 18–28.

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience, 5,* 629–630.

Schmidt, L., Belopolski, A. V., & Theeuwes, J. (2012). The presence of threat affects saccade trajectories. *Visual Cognition, 20,* 284–299. doi:10.1080/13506285.2012.658885

Sulman & Sanocki. (2011). *Top-down attentional capture by associated scenes in an object search task*. Paper presented at the 11th Annual Meeting of the Vision Science Society, Naples, FL.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381,* 520–522. doi:10.1038/381520a0

Thorpe, S., & Imbert, M. (1989). Biological constraints on connectionist models. In R. Pfeifer, Z. Schreter, F. Fogelman-Soulie, & L. Steels (Eds.), *Connectionism in perspective* (pp. 63–92). Amsterdam, the Netherlands: Elsevier.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136. doi:10.1016/0010-0285(80)90005-5

Vanrullen, R. (2009). Binding hardwired vs. on-demand feature conjunctions. *Visual Cognition, 17,* 103–119. doi:10.1080/13506280802196451

Vogt, J., De Houwer, J., & Moors, A. (2011). Unintended allocation of spatial attention to goal-relevant but not to goal-related events. *Social Psychology, 42,* 48–55. doi:10.1027/1864-9335/a000042

Vogt, J., De Houwer, J., Moors, A., Van Damme, S., & Crombez, G. (2010). The automatic orienting of attention to goal-relevant stimuli. *Acta Psychologica, 134,* 61–69. doi:10.1016/j.actpsy.2009.12.006

Wyble, B., Bowman, H., & Nieuwenstein, M. (2009). The attentional blink provides episodic distinctiveness: Sparing at a cost. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 324–37. doi:10.1037/a0013903

Wyble, B., Bowman, H., & Potter, M. (2009). Categorically defined targets trigger spatiotemporal attention. *Journal of Experimental Psychology: Human Perception and Performance 35* 787–807. doi:10.1037/a0013902

(*Appendix follows*)

Appendix

Category Names and Names of Target Pictures Used in Experiments 1, 1a, 2, and 3

Target category	First target	Second target
Amusement ride	rollercoaster	carousel
	bumper cars	ferris wheel
Art supply	crayons	paint brush
	paint	color pencils
Baby product	crib	bottle
	diaper	pacifier
Bathroom utility	toilet paper	toilet
	sink	bath tub
Bird	penguin	duck
	parrot	swan
	chicken	ostrich
	pigeon	peacock
Body part	feet/toes	fingernails
	hands	ear
	lips	nose
	tongue	eye
Carrying item	briefcase/suitcase	basket
	purse/handbag	backpack
Cleaning product	feather duster	broom
	mop	scrub/sponge
Computer part	mouse	printer
	monitor	keyboard
Dessert	chocolate cake	ice cream
	cookies	pie
Dinner food	hot dog	pizza
	hamburger	noodles/pasta
Flower	orchid	rose
	tulips	sunflower
Footwear	flip flops/sandals	boots
	running shoes	heels
Four-footed animal	bear	cow
	elephant	moose
	lion	ram/goat
	rhinoceros	dog
	horse	panda
	cat	pig
	giraffe	zebra
Fruit	bananas	papaya
	pear	strawberries
	watermelon	apple
	grapes	orange
	pineapple	pomegranate
Furniture	sofa	bed
	chair	chest of drawers
Gardening tool	rake	hedge scissors
	wheelbarrow	watering can
Insect	caterpiller	ladybird
Marine animal	sea horse	crab
	seal	octopus
	lobster	killer whale
	dolphin	walrus
Musical instrument	guitar	piano
	drums	viola/violin

(*Appendix continues*)

Appendix (*continued*)

Target category	First target	Second target
Personal hygiene art.	soap	nail cutter
	hair brush	toothpaste
	razor	tooth brush
Reptile	crocodile/alligator	snake
	tortoise	lizard
Sports equipment	hockey net	tennis ball
	basketball	baseball glove
Sports equipment	basketball net	tennis racket
	table tennis racquet	baseball
	soccer ball	ping pong table
Tableware	glass/cup	bowl
	fork	napkin
Tools	screw driver	hammer
Toy	rocking horse	teddy bears
	slinky	lego
Vegetable	cabbage	onion
	carrots	peas
	tomatoes	broccoli
	potatoes	peppers
	corn	onions
Vehicle	airplane	motorcycle
	truck	helicopter
	car	bicycle
	ship/cruise ship	bus
Weapon	knife	gun
	cannon	sword

Received June 1, 2012
Revision received July 5, 2012
Accepted July 12, 2012 ■