Picture Detection in RSVP: Features or Identity?

Mary C. Potter,
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology.

Brad Wyble,
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology.; Syracuse University, Syracuse, New York.

Rijuta Pandav, and
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology.

Jennifer Olejarczyk
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology.

Abstract

A pictured object can be readily detected in an RSVP sequence when the target is specified by a superordinate category name such as *animal* or *vehicle*. Are category features the initial basis for detection, with identification of the specific object occurring in a second stage (Evans & Treisman, 2005), or is identification of the object the basis for detection? When two targets in the same superordinate category are presented successively (lag 1), only the identification-first hypothesis predicts lag 1 sparing of the second target. The results of two experiments with novel pictures and a wide range of categories supported the identification-first hypothesis and a transient-attention model of lag 1 sparing and the attentional blink (Wyble, Bowman, & Potter, 2009).

Keywords

attentional blink; lag 1 sparing; picture processing; categorical search; transient attention

Studies in which viewers search for a target such as a digit in a stream of letters or an uppercase word among lowercase words show that detection is easy even when the items are presented as rapidly as 10/s (e.g., Chun & Potter, 1995; Lawrence, 1971). Pictured objects can also be detected readily: When viewing rapid serial visual presentation (RSVP) of a sequence of pictures with a rate of presentation as high as 9 pictures/s, participants can pick out a target picture described by a title such as "two men talking," (Potter, 1975, 1976; see also Intraub, 1980, 1981). With the presentation of a single picture for 20 ms, Thorpe and his colleagues (e.g., Rousselet, Fabre-Thorpe, & Thorpe, 2002) have shown that a viewer can readily detect whether the picture includes an animal (or, in other experiments, a means of transportation). Measures of evoked potentials on the scalp begin to differentiate target-present from target-absent trials as early as 150 ms after the picture is onset (Thorpe, Fize, & Marlot, 1996). When the target is one of two pictures presented simultaneously, eye

movements to the target picture are initiated as early as 120 ms after onset (Kirchner & Thorpe, 2006).

The basis for detection of pictured targets defined by a category or a title is not well-understood, however. Unlike letters, digits, or words, the pictures in these studies have never been seen before. Is the picture as a whole understood within about 100 ms, or is detection based on one or more features that are characteristic of the target category, as suggested by Evans and Treisman (2005)? They hypothesized that features or characteristic parts of objects in a given category (such features as beaks, claws, fur, or eyes that characterize animals) may be perceived in parallel early in processing, permitting detection of the category "animal." Consistent with Treisman's feature integration theory of attention (Treisman & Gelade, 1980), they proposed that additional serial processing is required to select and bind those features at a particular location and to set up an object file, allowing the viewer to become conscious of the object and report it.

The two stages of processing suggested by Evans and Treisman (2005) are similar to the two-stage model of Chun and Potter (1995) for detection of letters among digits, in that the first stage is fast and parallel[1], while the second stage is serial. Such models have been tested by presenting two successive targets (T1 and T2) at varying stimulus onset asynchronies (SOAs) in the RSVP stream. Both models predict that at short SOAs there will be a delay in processing T2 while T1 is being processed in the second stage, causing T2 to be missed--an attentional blink (Raymond, Shapiro, & Arnell, 1992; see also Broadbent & Broadbent, 1987, and Weichselgartner & Sperling, 1987). A critical difference between the Evans-Treisman and Chun-Potter models is the stage at which targets are identified. In Chun and Potter's model identification occurs in Stage 1; the loss of T2 results from a failure to consolidate its memory in Stage 2. In the Evans-Treisman model of picture detection, unbound features are detected in Stage 1, but identification of the object occurs only after feature-binding and the setting up of an object file, in Stage 2; crucially, only one object can be bound at a time. As discussed below, the two models make different predictions about what happens when the two targets occur in immediate succession, at lag 1.

Lag 1 sparing and transient attention

In many search tasks with familiar categories such as digits, letters, and words there is a reduction or elimination of the attentional blink when the second target appears within an SOA of about 100 ms, so-called lag 1 sparing (Potter, Chun, Banks, & Muckenhoupt, 1998; see Visser, Bischof, & Di Lollo, 1999, for a review). It has been proposed that this surprising reduction in the attentional blink occurs because the first target opens an attentional gate long enough to let in an immediately following target, and the two targets are processed together in the second stage (e.g., Chun & Potter, 1995). Along similar lines, another explanation of lag 1 sparing (Wyble, Bowman, & Potter, 2009) proposes that detection of a target elicits a transient burst of attention that will facilitate detection of a following target, but only if the latter appears within a window of about 150 ms after the onset of the first target.

In early studies of transient attention (Muller & Rabbitt, 1989; Nakayama & Mackeben. 1989) the stimulus that generated transient attention at a given location was a visually salient cue. In Wyble, Bowman, and Potter's study (2009), however, the cue attracted attention because it was itself a target, not because it was visually salient. The search array consisted of eight changing items in different locations; the target could occur in any one of the

---

[1]The Chun-Potter model was not specific about whether processing in Stage 1 is parallel or simply fast enough to be completed within an exposure duration as short as 100 ms

locations. Two targets were presented successively, either in the same location or in different locations. The targets were defined by category (e.g., digits among keyboard symbols) rather than by visual salience. As in earlier studies of transient attention, the cue (the leading target) only benefited the critical target when they appeared in the same location and the critical target was presented within 150 ms after the onset of the leading target.

Wyble, Bowman, and Potter (2009) suggested a link between the transient attention shown in their study and lag 1 sparing in studies of the attentional blink. Even when all stimuli appear in a single location, transient attention from the first target would enhance processing of a T2 that arrives within a 150 ms temporal window after T1. In this model, which is described more explicitly in Wyble, Bowman, and Nieuwenstein (2009). transient attention is suppressed about 150 ms after target onset, to prevent it from being triggered again while T1 is being processed into working memory. Suppression of transient attention enhances the episodic information inherent in the visual stream, by separating the encoding of T1 and T2 representations. It is this suppression of transient attention that causes the attentional blink at SOAs longer than 150 ms. The model is similar to that described by Chun & Potter (1995) in that it is a two stage model which can encode multiple items within a single attentional window.

Several additional observations support the hypothesis that lag 1 sparing is the result of transient attention. Like transient attention, lag 1 sparing typically occurs only if T1 and T2 are presented in the same location (Visser et al., 1999). Consistent with transient attention (but not with other explanations of lag 1 sparing) there is sparing even with an intervening distractor between T1 and T2 (so that T2 is actually at lag 2), as long as the SOA between the targets is less than 150 ms (Bowman & Wyble, 2007; Potter, Staub, & O'Connor, 2002; Wyble, Bowman, & Potter, 2009).

Differential predictions about lag 1 sparing in picture search

In Evans and Treisman's model (2005), a second target at lag 1 arrives before T1 has been identified; at that point, only target features have been detected and have not been bound into an object file. Transient attention from detection of the features of the first target could benefit detection of features of the second target at lag 1, possibly importing both sets of features into Stage 2. Because Stage 2 can only bind one object at a time, however, one or the other object might be identified, but not both. Alternatively, if Stage 2 processing of the first target is initiated before the next picture appears (cutting off the next picture's access to Stage 2), there should be a marked attentional blink for the second target at lag 1, worse than that at lag 2. In neither case should there be lag 1 sparing on trials in which T1 is reported. Lag 1, however, was not included in Evans and Treisman's study or in any previous study with pictures.[3]

The transient attention model of Wyble, Bowman, and Potter (2009), like the Chun-Potter (1995) two-stage model, proposes that categorical targets are detected in Stage 1 as they are identified. Transient attention from detection of the first target overlaps with the following target, increasing the probability that it will also be identified and detected (lag 1 sparing).

To distinguish between the Evans-Treisman and the Wyble-Bowman-Potter models, in the present study we ask whether there is lag 1 sparing for picture targets when the particular picture has never been seen before and only its superordinate category has been specified. To reduce the likelihood that participants would develop shortcuts that would enable them to

---

[3]An exception is an attentional blink study by Dux and Harris (2007) that used line drawings of objects. They did not observe lag 1 sparing; we have no immediate explanation for this difference between line drawings and color photographs.

detect targets without specifically identifying them, we used a wide range of categories and exemplars, we informed the participant about the category on a given trial only seconds before the pictures appeared, and (like Evans & Treisman, 2005) we required the participant to report the specific identity (the basic level category) of each of the targets.

Experiment 1

In Experiment 1 participants searched for two targets in an RSVP sequence of eight pictures presented at the rate of 107 ms/picture. The targets were pictures of single objects that were specified by a superordinate category such as "vehicle," "fruit," etc.; both targets on a trial were from the same category. Examples of the target pairs are shown in Figure 1. The distractors were pictures of single objects or scenes. The second target was presented at lag 1, 2, or 4 (stimulus onset asynchronies--SOAs--of 107, 213, and 427 ms, respectively).

Method

Participants—Twenty-four subjects (11 male; 1 left-handed) from the M.I.T. community participated in the experiment and were paid. All participants had normal or corrected to normal vision and none reported being color-blind. They were all native speakers of American English. One other subject was replaced because his accuracy was more than two standard deviations below that of the other subjects.

Apparatus—The experiment was run using Matlab 5.2.1 (build 1421) on a G3 . The screen was set to 1024×768 resolution at a 75 Hz refresh rate on an Apple 17" studio display.

Stimuli—The stimuli consisted of colored photographs of single objects in their natural settings and pictured scenes. The pictures were downloaded from Google Images. There were a total of 160 single object target pictures, 160 distractor object pictures and 320 distractor scenes. 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. The pictures were presented in the center of the screen on a gray background; the horizontal visual angle was 10.3°, at the normal viewing distance of 50 cm.

Pairs of target pictures were selected from 29 superordinate categories of objects such as fruit, vehicle, body part, or cleaning product; the pictured exemplars were typical of their basic level categories, e.g., banana, boat, ear, or broom. The number of pairs per category ranged from 1 (insect) to 8 (four-footed animal). The pairs were chosen to avoid closely related or similar-looking exemplars of the category. The appendix gives a list of the categories and target pairs. The 160 distractor pictures of single objects were from superordinate categories other than the target categories. The 320 distractor scenes included more than one object, for example a woman drinking tea, people in a factory, fireplace.

Design and procedure—Each trial consisted of an RSVP stream of eight color photographs with two target pictures and six distractor pictures (two pictures of objects and four scenes). The first target (T1) appeared in serial position two or three, counterbalanced within and between subjects. The second target (T2) followed the first at lag 1 (SOA 107 ms), 2 (SOA 213 ms) or 4 (SOA 427 ms), counterbalanced within and between subjects. A trial is illustrated in Figure 2. Which of the two targets in a given trial appeared first was counterbalanced between subjects.

All subjects saw 8 practice trials and 72 experimental trials. Each trial began with a fixation cross for 400 ms followed by the category name of the targets, presented in black size 20 Courier font and displayed for 750 ms. The fixation cross reappeared for 300 ms, followed by a blank interval of 300 ms and then the picture sequence; each picture was presented for

107 ms. At the end of the sequence there was a blank screen for 107 ms, followed by a dialog box with spaces for entering two responses; subjects were encouraged but not required to enter the object names in the order of presentation. They were instructed to give the name of the object, not its category, and if they did not know the name, to describe the object. Subjects pressed the space bar to begin and end each trial, proceeding at their own pace. Most subjects completed the experiment within 30 minutes.

Scoring—The written responses were scored as correct if they were the name we gave the object or a synonym for that name. Responses that were the name of a closely related object in the same category for which the object might have been mistaken such as papaya and mango were also counted as correct (3.4% of correct responses), as were responses that provided a close, correct description of the object (less than 1% of correct responses). All other responses and omissions were scored as incorrect. The order of responses was ignored.

Results and Discussion

The main results are shown in Figure 3. Separate analyses of variance (ANOVAs) were carried out on the proportion of correct responses to T1 and to T2 conditional on a correct response to T1. Variables were the serial position of T1 (second or third picture) and the SOA between T1 and T2. In the analysis of T1 accuracy there was no effect of the serial position of T1, but a significant effect of SOA, with a lower performance at lag 1 (M = .74) than at lags 2 or 4 (both Ms = .84), $F(2, 46) = 4.47$, $p < .02$, $\eta_p^2 = .16$. Competition between T1 and T2 at lag 1 is likely to be responsible for the lower performance of T1 at lag 1, a pattern frequently observed in attentional blink studies. There was no interaction with serial position.

In the analysis of T2 accuracy conditional on a correct T1, the serial position of T1 had no effect, and did not interact with SOA. There was a main effect of SOA, $F(2, 46) = 20.449$, $p < .001$, $\eta_p^2 = .47$, with lag 1 sparing (M = .72), an attentional blink at lag 2 (M = .52), and recovery from the blink at lag 4 (M = .78). Thus, the hypothesis that detection of T1 would trigger transient attention, leading to lag 1 sparing, was supported by the results.[2]

These results are inconsistent with the Evans-Treisman (2005) model of categorical detection based on unbound features, because transient attention generated by detection would lead to confusion among two sets of unbound features, rather than to lag 1 sparing. Instead, the results support a model in which the basic level category of a pictured object is identified at or before the moment that it is detected as belonging to the superordinate target category, enabling two targets in close succession to be identified in Stage 1 and then consolidated together in Stage 2. The time course of attentional deployment in response to a novel categorically defined picture target appears to be similar to that observed in RSVP paradigms using simpler familiar stimuli, such as letters and digits. This finding fits well with recent research which shows rapid behavioral responses to pictures of a specified target category (e.g., Kirchner & Thorpe, 2006) and is consistent with research showing that electrophysiological correlates of target detection begin 150 ms after target onset (Thorpe et al., 1996).

If lag 1 sparing of the second target is truly the result of a transient deployment of attention to the first target, the first target should capture attention to its location. In experiments using multiple spatial locations and categorically defined targets, such as letters and digits presented amidst symbol distractors, report of a second target was found to depend critically

[2]A separate analysis of T2 accuracy conditional on not reporting T1, with SOA as the only variable, showed a significant effect of SOA, $F(2, 42) = 8.697$, $p < .001$, with correct responses on .86, .72, and .58 for lags of 1, 2, and 4, respectively (two subjects were omitted because they did not miss any T1s in one or more of the conditions).

on its spatial and temporal proximity to a first target (Wyble, Bowman, & Potter, 2009). Trials in which T2 appeared in the same location as T1 at an SOA of about 100 ms showed a marked increase in report of T2 (similar to the location specificity of lag 1 sparing in other AB tasks, Visser et al., 1999), compared to control trials without T1. In contrast, when T1 was in a different location, report of T2 was impaired, relative to the no-T1 condition, suggesting that attention was captured to the location of T1. In Experiment 2 we tested the effect of the relative spatial locations of T1 and T2 on the lag 1 sparing we found in Experiment 1.

Experiment 2

Experiment 2 examined lag 1 performance for a second target in the same or a different spatial location as the first target. If lag 1 sparing is due to transient attention initiated by the first target, then one would expect it to be restricted to the location of the first target; in fact, there should be interference with T2 processing if T1 appears in a different location. Participants searched for one or two targets in two adjacent RSVP streams. As in Experiment 1, targets were pictures of single objects specified by a category such as vehicle or fruit. The method differed from Experiment 1 in that there were only three frames in the RSVP sequence, each with two pictures side by side, as illustrated in Figure 4. T1 plus a distractor (or just two distractors on no-T1 trials) appeared in the first frame, T2 plus a distractor in the second frame, and two non-target pictures in the third frame. Thus, in all the trials with two targets T2 was presented at lag 1 (at an SOA of 80 ms) relative to T1. Critically, on half the trials with a T1, T2 was in the same location (on the same side) as T1, and half in the other location.

Method

The method was the same as that of Experiment 1, except as noted.

Participants—Twenty-five members of the M.I.T. community (11 males) participated in the experiment; none had participated in Experiment 1. One participant was replaced because his performance was significantly lower than that of the other subjects.

Apparatus and stimuli—The apparatus was the same as that in Experiment 1, as were the target stimuli and category names.[4] Distractors were the single-object distractors used in Experiment 1, together with 66 additional pictures of single objects obtained from the web. Pictured scenes taken from the set in Experiment 1 were used as masks following the T2 frame. The pictures were presented on either side of the fixation cross, on a gray background. The visual angle of the pair of pictures, at the normal viewing distance of 50 cm, was 21.9°.

Design and procedure—The RSVP streams consisted of 3 frames, each with two pictures on either side of a central fixation cross. On 2/3 of the trials the first frame included T1 plus a single-object distractor. Half the time T1 was on the left, half on the right. On 1/3 of the trials T1 was replaced by a single-object distractor. The second frame always included T2 alongside a single-object distractor. When T1 had been presented, half the time T2 was on the same side, counterbalanced for left-right location. The third frame consisted of pictured scenes that functioned as masks. In comparison with Experiment 1, in which all frames appeared for 107 ms, in Experiment 2 the duration of the first frame was shortened to 80 ms because of the extra attention commanded by the first frame in a sequence, and also to

---

[4]The 8 practice trials in Experiment 1 became regular trials in Experiment 2, and 8 regular trials in Experiment 1 became the practice trials in Experiment 2; see the Appendix for details.

produce enough T1 misses to permit comparisons between three conditions: T1 present and reported, T1 present but not reported, and T1 omitted. The duration of the second frame was increased to 120 ms because of the increased difficulty of processing two simultaneous pictures (cf. Potter & Fox, 2009). Thus, the duration of the first frame was 80 ms; the second, 120 ms; and the third, 107 ms. Which of the two targets in a given pair appeared as T1 was counterbalanced between subjects.

All participants had 8 practice trials and 72 experimental trials. Each trial began with a fixation cross for 400 ms, followed by the category name of the targets, displayed for 750 ms. After another fixation cross for 300 ms, and a blank screen for 300 ms, the RSVP sequence appeared, followed by a blank screen of 50 ms and then a dialog box with spaces for entering responses. Subjects pressed the space bar to begin and end each trial; the experiment took about 30 minutes to complete.

Scoring—Responses were scored for correctness as in Experiment 1.

Results and Discussion

The main results are shown in Figure 5. Consistent with the transient attention hypothesis, T2 was more likely to be reported when T1 appeared in the same location as T2 than when T1 appeared in the other location. Analyses were carried out separately on T1 and T2 performance. In the analysis of T1, the effect of whether T1 was presented in the same location as T2 was not significant, $p = .23$, although there was a tendency for lower accuracy when T1 was in the same location (M = .52 versus .56), consistent with competition from T2 at that location. As expected because of the shortened duration of T1 and the simultaneous presentation of a distractor with T1, accurate report of T1 was lower in Experiment 2 (.54) than in Experiment 1 (.81).

In the analysis of correct responses to T2 in all three conditions (whether or not T1 was reported), the effect of condition was significant, as shown in Figure 5a, $F(2, 46) = 6.35$, Mse = 0.0147, $p < .01$, $\eta_p^2 = .22$. Report of T2 alone was not different from T2 preceded by T1 in the same location, whereas report of T2 preceded by T1 in a different location was lower than the other two conditions. In a further analysis of T2 performance conditional on a correct T1 (in the two conditions in which T1 had been presented), T2 on the same side as T1 was significantly better (M = .58) than on the other side (M = .45), $F(1, 23) = 8.025$, Mse = 0.0250, $p < .01$, $\eta_p^2 = .26$, consistent with the hypothesis that lag 1 sparing is the result of location-specific transient attention generated by T1.

To examine the effect of reporting or missing T1 on report of T2, we carried out an analysis of T2 with report of T1 (correct versus incorrect) and same-different location as the variables (see Figure 5b). The effect of same versus other location was again significant, $F(1, 23) = 9.504$, Mse = 0.0190, $p < .01$, $\eta_p^2 = .29$. Performance on T2 was better when T1 had been missed, $F(1, 23) = 14.933$, Mse = 0.0439, $p = .001$, $\eta_p^2 = .39$, indicating that there was some competition between T1 and T2 that was reduced when T1 was not reported. Although the interaction with location was not significant, $p = .20$, the benefit of missing T1 was somewhat greater (.21 versus .12) when the targets were in different locations, suggesting that it was particularly difficult to see T2 if T1 was seen and reported in a different location. Comparing performance on T2 when T1 had not been presented (Figure 5a, M = .65) with performance when T1 was presented in the same location but was not reported (Figure 5b, M = .70) suggests that a missed T1 produced some degree of transient attention that facilitated T2 in that location; a similar tendency was seen in Experiment 1 at lag 1 (Footnote 2).

The results of Experiment 2 show that T1 attracts attention to its location, so that T2 performance is better in that location than in the other location. This result is consistent with the transient attention hypothesis in showing that the attention deployed by T1 occurred rapidly enough to affect the accuracy of a following T2, even though T1 appeared only 80 ms before T2 and frequently could not be reported.

General Discussion

Experiment 1, which used a single RSVP stream, demonstrated that novel pictured targets specified by category show lag 1 sparing of the second target, indicating that transient attention elicited by detection of the first target benefits a second target arriving immediately after the first. By an SOA of 213 ms (lag 2), sparing was replaced by an attentional blink for T2, and by an SOA of 427 ms the blink was over and T2 was reported almost as accurately as T1. Experiment 2, with two streams of pictures, showed that lag 1 sparing was reduced when T2 was presented at a different location than T1, consistent with previous studies of transient attention (e.g., Nakayama & Mackeben, 1989).

Evans and Treisman (2005) found that viewers can detect pictures of animals or vehicles, presented in an RSVP stream for 110 ms per picture. They proposed that participants detected pictures by parallel activation of characteristic features such as fur or a beak, and that identification of the specific object occurred in a second, serial stage in which features at a given location were bound, resulting in conscious identification that made it possible to report the target. During the serial stage of T1 no binding of a second target picture could occur, consistent with the substantial attentional blink that they obtained at lag 2. They did not include lag 1 in their study, but their theory states explicitly that no more than one object can be bound and identified at a time. Thus, their theory appears to rule out the possibility of processing both T1 and T2 at lag 1; at best, the features of both objects should be intermixed, presenting a sorting-out problem in Stage 2 that would produce many errors of identification, and no lag 1 sparing.

A further problem with the Evans-Treisman theory is that the features that characterize a category and allow detection must be fairly general to objects in that category, whereas the task in the present experiments (as in the Evans-Treisman experiments) was to report the specific identity of each target: otherwise, responses were not counted as correct. Beaks, wings, and claws indicate a bird, but other features are needed to identify it as a penguin or a duck. That is, the features that distinguish individual exemplars would not match "the active nodes in the recognition network" (Evans & Treisman, 2005, p. 1490) that are primed by the category name given at the beginning of the trial. Only after the features at a given object location are bound will the object's specific identity be known.

In the present experiments T1 was correctly reported on a majority of lag 1, same position trials in both experiments (.74 and .52, respectively), and T2 was correctly reported along with T1 on .71 and .58 (respectively) of those trials. Moreover, in Experiment 1 performance on T2 was markedly better at lag 1 than at lag 2, which showed an attentional blink. These results cast the Evans-Treisman features explanation of rapid picture detection into doubt. Instead, the results support the hypothesis that a pictured object is not only detected but also identified as a specific instance of the target category within about 100 ms, generating a burst of transient attention that facilitates processing of the immediately following object (within about 150 ms of the onset of T1). If T2 appears at lag 2 (an SOA of 213 ms), that is too late to benefit from transient attention and T2 is subject to an attentional blink. In Wyble, Bowman, and Potter's model (2009), the blink is the result of inhibition of the attentional blaster (transient attention) while the current stimulus is being encoded into short term memory as a distinct episode. That is, both sparing at lag 1 and an attentional blink at longer lags result from the operation of transient attention.

Detection in the present experiments could not be based on visual salience, but had to be based on categorical membership. As the categories included diverse exemplars, it would have been difficult for the participant to anticipate specific exemplars and features of the target, particularly as the category on a given trial was specified less than 1.5 s before the sequence began and the categories were of such broad scope that specific colors or textures would not have been particularly useful marks of category membership. Thus, objects presumably had to be individually identified, perhaps in a feedforward manner, in order to be categorized. VanRullen (2009) suggests that familiar objects have hardwired binding of their features that allows them to be identified without selective attention, unlike arbitrary, ad hoc stimuli such as colored geometric shapes that do require on-demand attention to be bound. Finally, it should be mentioned that the human ability to rapidly identify objects is consistent with recent neurophysiological work in monkeys (e.g., Hung, Kreiman, Poggio, & DiCarlo, 2005), which shows that identity-specific representations are activated in inferotemporal regions within 125 ms of stimulus onset.

In conclusion, the present experiments indicate that a pictured object can be identified as belonging to a target category in time for a following target object also to be selected, encoded, and reported, when targets are presented as briefly as 80-120 ms/picture. This result shows that such rapid identification is not restricted to stereotypical visual categories such as letters, digits, or even words, but holds also for diverse categories of familiar objects presented in pictures the observer has never seen before.

Acknowledgments

Appendix: Materials in Experiments 1 and 2

Category names and names of target pictures used in Experiments 1 and 2. Half the subjects in each experiment saw the picture targets in the first-second order shown; the other half saw them in the reverse order.

Target Category	First Target	Second Target
Amusement ride	rollercoaster	carousel
	[a]bumper cars	ferris wheel
Art supply	crayons	paint brush
	paint	color pencils
Baby product	crib	bottle
	[a]diaper	pacifier
Bathroom utility	toilet paper	toilet
	sink	bath tub
Bird	[a]penguin	duck
	parrot	swan
	chicken	ostrich
	pigeon	peacock
Body part	[b]feet/toes	fingernails[c]
	hands	ear
	lips[d]	nose
	tongue	eye

Target Category	First Target	Second Target
Carrying item	briefcase/suitcase	basket
	[b]purse/handbag	backpack
Cleaning product	feather duster	broom
	[b]mop	scrub/sponge
Computer part	mouse	printer
	monitor	keyboard
Dessert	chocolate cake	ice cream
	[a]cookies	pie
Dinner food	hot dog	pizza
	hamburger	noodles/pasta
Flower	[b]orchid	rose
	tulips	sunflower
Footwear	[b]flip flops/sandals	boots
	[a]running shoes	heels
Four-footed animal	bear	cow
	[a]elephant	moose
	lion	ram/goat
	rhinoceros	dog
	horse	panda
	kitten	pig
	giraffe	zebra
	sheep/llama	tiger
Fruit	[a]banana	papaya
	pear	strawberries
	watermelon	apple
	grapes	orange
	pineapple	pomegranate
Furniture	sofa	bed
	chair	chest of drawers
Gardening tool	rake	hedge scissors
	[b]wheelbarrow	watering can
Insect	caterpiller	ladybird
Marine animal	sea horse	crab
	seal	octopus
	[a]lobster	killer whale
	dolphin	walrus
Musical instrument	guitar	piano
	drums	viola/violin
Personal hygiene art.	soap	nail cutter
	hair brush	toothpaste
	razor	tooth brush
Reptile	crocodile/alligator	snake

Target Category	First Target	Second Target
	tortoise	lizard
Sports equipment	hockey net	tennis ball
	basketball	baseball glove
	[a]basketball net	tennis racket
	[b]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	garlic
	carrots	peas
	tomatoes	broccoli
	potatoes	peppers
	corn	onions
Vehicle	airplane	motorcycle
	truck	helicopter
	car	bicycle
	ship/cruise ship	bus
Weapon	knife	gun
	cannon	sword

[a]practice pair in Experiment 2

[b]practice pair in Experiment 1

[c]changed to *lips* in Experiment 2

[d]changed to *fingernails* in Experiment 2

References

Broadbent DE, Broadbent MH. From detection to identification : Response to multiple targets in rapid serial visual presentation. Perception & Psychophysics 1987;42:105–113. [PubMed: 3627930]

Chun MM, Potter MC. A two-stage model for multiple target detection in rapid serial visual presentation. Journal of Experimental Psychology: Human Perception and Performance 1995;21:109–127. [PubMed: 7707027]

Dux PE, Harris IM. Viewpoint costs occur during consolidation: Evidence from the attentional blink. Cognition 2007;104:47–58. [PubMed: 16797522]

Evans KK, Treisman A. Perception of objects in natural scenes: Is it really attention-free? Journal of Experimental Psychology: Human perception and Performance 2005;31:1476–1492. [PubMed: 16366803]

Hung CP, Kreiman G, Poggio T, DiCarlo JJ. Fast readout of object identity from macaque inferior temporal cortex. Science Nov 4;2005 310:863–866. [PubMed: 16272124]

Intraub H. Presentation rate and the representation of briefly glimpsed pictures in memory. Journal of Experimental Psychology: Human Learning and Memory 1980;6:1–12. [PubMed: 7373241]

Intraub H. Rapid conceptual identification of sequentially presented pictures. Journal of Experimental Psychology: Human Perception and Performance 1981;7:604–610.

Kirchner H, Thorpe SJ. Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. Vision Research 2006;46:1762–1776. [PubMed: 16289663]

Lawrence DH. Two studies of visual search for word targets with controlled rates of presentation. Perception & Psychophysics 1971;10:85–89.

Muller HJ, Rabbitt PM. Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. Journal of Experimental Psychology: Human Perception and Performance 1989;15(2):315–330. [PubMed: 2525601]

Nakayama K, Mackeben M. Sustained and transient components of focal visual attention. Vision Research 1989;29(11):1631–1647. [PubMed: 2635486]

Potter MC. Meaning in visual search. Science 1975;187:965–966. [PubMed: 1145183]

Potter MC. Short-term conceptual memory for pictures. Journal of Experimental Psychology: Human Learning and Memory 1976;2:509–522. [PubMed: 1003124]

Potter MC, Chun MM, Banks BS, Muckenhoupt M. Two attentional deficits in serial target search: The visual attentional blink and an amodal task-switch deficit. Journal of Experimental Psychology: Learning, Memory, and Cognition 1998;24:979–992.

Potter MC, Fox LF. Detecting and remembering simultaneous pictures in a rapid serial visual presentation. Journal of Experimental Psychology: Human Perception and Performance 2009;35:28–38. [PubMed: 19170468]

Potter MC, Levy EI. Recognition memory for a rapid sequence of pictures. Journal of Experimental Psychology 1969;81:10–15. [PubMed: 5812164]

Potter MC, Staub A, O'Connor DH. The time course of competition for attention: Attention is initially labile. Journal of Experimental Psychology: Human Perception and Performance 2002;28:1149–1162. [PubMed: 12421061]

Raymond JE, Shapiro KL, Arnell KM. Temporary suppression of visual processing in an RSVP task: An attentional blink? Journal of Experimental Psychology: Human Perception and Performance 1992;18:849–860. [PubMed: 1500880]

Rousselet G, Fabre-Thorpe M, Thorpe SJ. Parallel processing in high level categorization of natural images. Nature Neuroscience 2002;5:629–630.

Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. Nature June 6;1996 381:520–522. [PubMed: 8632824]

VanRullen R. Binding hardwired versus on-demand feature conjunctions. Visual Cognition 2009;17:103–119.

Visser TAW, Bischof WF, Di Lollo V. Attention switching in spatial and nonspatial domains: Evidence from the attentional blink. Psychological Bulletin 1999;125:458–469.

Weichselgartner E, Sperling G. Dynamics of automatic and controlled visual attention. Science 1987;238:778–780. [PubMed: 3672124]

Wyble B, Bowman H, Potter MC. Categorically defined targets trigger spatiotemporal attention. Journal of Experimental Psychology: Human Perception and Performance 2009;35:324–337. [PubMed: 19331491]

Wyble B, Bowman H, Nieuwenstein M. The attentional blink provides episodic distinctiveness: Sparing at a cost. Journal of Experimental Psychology: Human Perception and Performance 2009;35:787–807. [PubMed: 19485692]
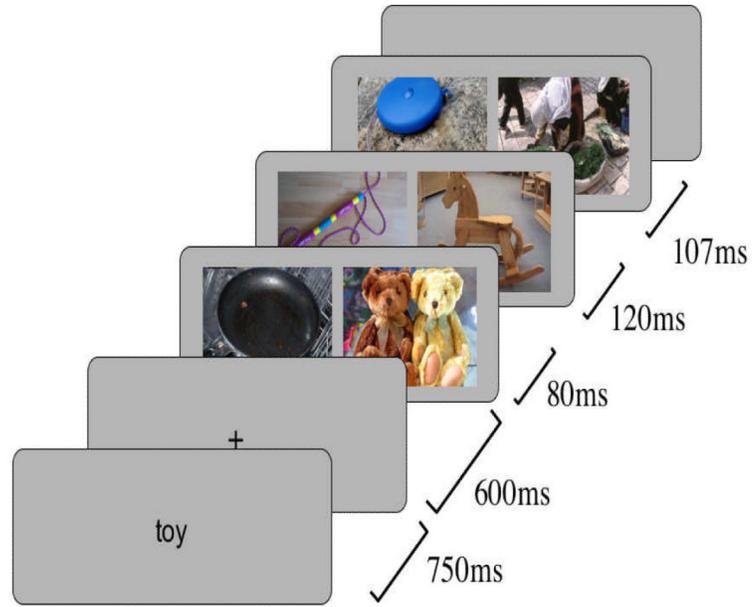
Figure 1.
Examples of pairs of targets in given categories.
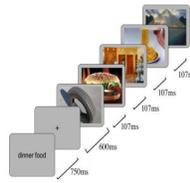
Figure 2.
An illustration of a trial in Experiment 1. Only part of the eight-picture sequence is shown; the targets are *hamburger* and *noodles/pasta*.
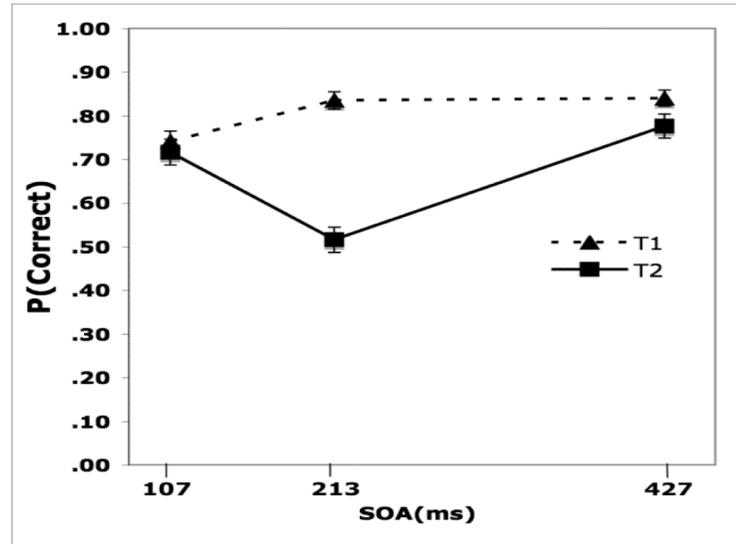
Figure 3.
Experiment 1: Proportion of correct responses at each SOA for T1 (dashed line) and T2 conditional on a correct T1 (solid line). Error bars are standard errors of means.

Figure 4.
An illustration of a trial in Experiment 2, with T1 (*teddy bears*) and T2 (*rocking horse*) in the same location.
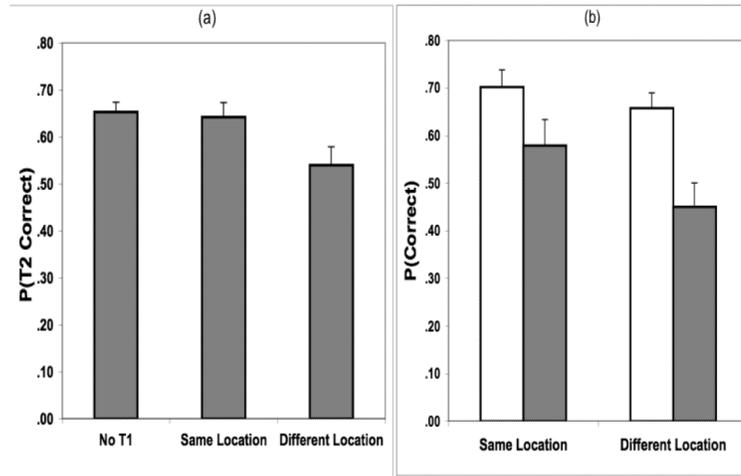
Figure 5.
Experiment 2: (a) Proportion of correct responses to T2 when there was no T1, when T1 was in the same location as T2, and when T1 was in a different location than T2. (b) Proportion of correct responses to T2 conditional on a correct response to T1 (gray bars) or conditional on a wrong response to T1 (white bars), when T1 and T2 were presented at the same location or different locations. Error bars are standard errors of means.