The attentional blink

Kimron L. Shapiro, Karen M. Arnell and Jane E. Raymond

When two masked targets (T1 and T2) are presented within approximately 500 ms of each other, subjects are often unable to report the second of the two targets (T2) accurately, even though the first has been reported correctly. In contrast, subjects can report T2 accurately when instructed to ignore T1, or when T1 and T2 are separated by more than 500 ms. The above pattern of results has been labelled the attentional blink (AB). Experiments have revealed that the AB is not the result of perceptual, memory or response output limitations. In general, the various theories advanced to account for the AB, although they differ in the specific mechanisms purported to be responsible, assume that allocating attention to T1 leaves less attention for T2, rendering T2 vulnerable to decay or substitution. The present report attempts to bring together these various accounts by proposing a unifying theory. This report also highlights recent attempts to determine if the AB exists across stimulus modalities and points to applications of AB methods in understanding deficits of visual neglect. We conclude by suggesting that investigations of the AB argue in favour of the view that attention may be thought of as a necessary (but not sufficient) condition for enabling consciousness.

Typically, humans have difficulty acquiring conscious knowledge of all stimuli arriving simultaneously from multiple sources. For example, while watching a sporting event, you may fail to notice that the person sitting beside you is no longer there, even though they walked in front of you as they departed. Cognitive scientists attribute such lapses in awareness of seemingly obvious stimuli to temporary losses in attention. Attention is taken to mean a set of integrated neural processes that selects a given perceptual input from among competing inputs. In this article, we argue that the further function of attention is to allow selected perceptual information a foothold in consciousness.

Limits governing the brain's ability to process sequentially presented stimuli can be studied using rapid serial visual presentation (RSVP). In RSVP methodology, stimuli such as letters, digits, words and pictures are presented briefly in the same location and in rapid succession, at rates of about 10 items per second with participants having to identify one or more of these stimuli (targets). Any single target, although presented for only 100 ms, can be reported accurately[1]. However, we and others before us, have shown that reporting the second of two consecutively presented targets successfully is considerably more difficult when these targets occur within a short interval of each other[2–5].

To study the timecourse of attention, the present authors developed a two-target RSVP procedure as shown and described in Fig. 1 (Ref. 3). We record the proportion of correct responses to the second target (T2), contingent on first target (T1) correct responses, as a function of the T1–T2 interval. The typical result (see Fig. 2) is that the proportion of correct T2 responses in the dual-target task falls dramatically in the first half second following T1 and recovers thereafter. We use the term 'attentional blink' (AB) to describe this outcome. It is important to note three conclusions drawn by Raymond *et al.*[3] which are still pertinent to this basic finding: (1) the AB cannot be explained by low-level sensory factors (such as retinal masking); (2) the AB cannot be attributed to memory span limitations; and (3) T1 processing must be interrupted by another visual stimulus or the AB effect will not occur.

The attentional blink is not like a real eyeblink

In 1992, Raymond *et al.*[3] suggested that the AB could be the result of T1 tying up attentional resources completely and leaving insufficient resources to process any aspects of T2. Recent experiments[6–10] have shown that we were correct in so far as T2 is not processed to a level sufficient for report but incorrect in our supposition that no aspects of T2 are processed. Two experiments revealing that T2 is processed to a stage just short of awareness are described below.

The rationale for the first experiment[6] stems from the masked priming literature where it has been shown that a stimulus, presented too briefly to be reported, nevertheless, can facilitate report of (or prime, as it is called) an identical or semantically related stimulus. For example, the word 'doctor', although not recognized, will prime a second presentation of 'doctor' or 'nurse' better than the word 'lawyer' when presented before either of these words. If in an AB experiment, T2 items presented during the AB interval can prime a subsequently presented target, we argued that this would demonstrate a high level of processing during the AB. Using a three target AB paradigm, Shapiro *et al.*[6] showed that when participants were unable to report T2 (that is, it was 'blinked'), nevertheless, this stimulus acted to

K.L. Shapi
J.E. Raymon
the School of Psychology, The University of Wales, Bangor, UK LL57 2DG.
K.M. Arnell is at the MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge, UK CB2 2EF.

tel: +44 1248 383626
fax: +44 1248 382599
e-mail: k.shapiro@bangor.ac.uk

 1364-6613/97/$17.00 PII: S1364-6613(97)01094-2

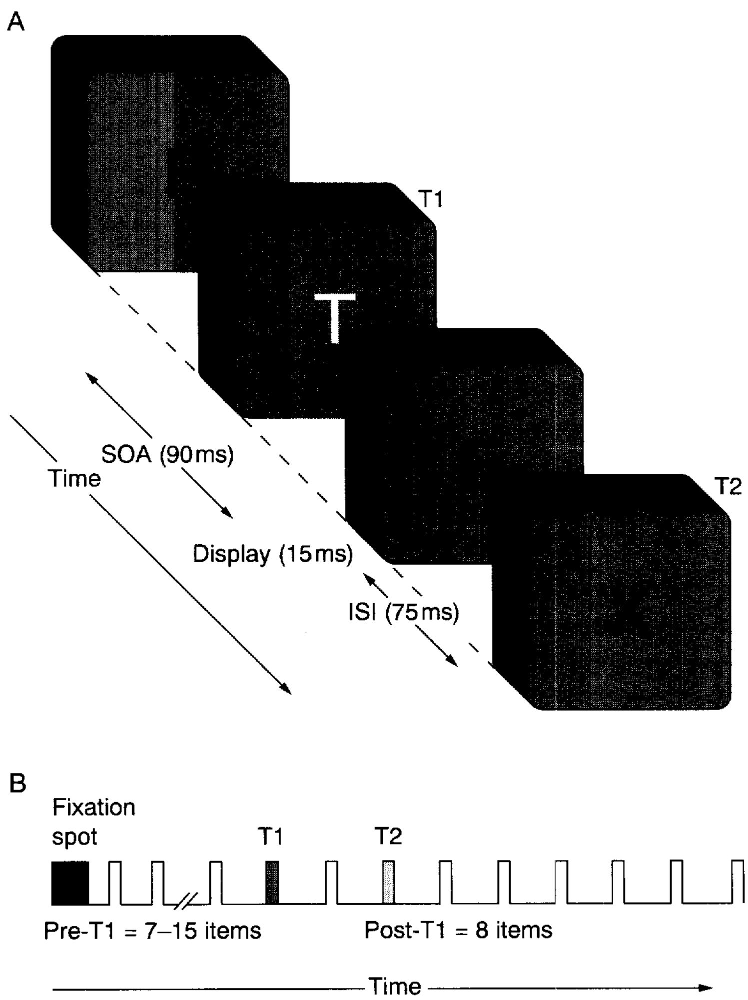


Fig. 1 The dual-target rapid serial visual presentation (RSVP) test. (A) Indicates the temporal parameters used in the particular RSVP procedure described in the text. **(B)** The dual-target (experimental) condition, depicted schematically, shows a variable number of black, randomly chosen letters preceding the first target (T1). The T1 task is to identify the white letter, exemplified as a 'T' in (A). In a randomly chosen half of the trials, T2 is presented in one of the eight possible temporal positions following T1, whereas in the remaining half of the trials no T2 is presented (the T1-absent trials allow false alarm rates to be determined). The T2 task is to detect if the letter 'X' is presented. Randomly selected filler letters are shown in the post-T1 positions not occupied by T2. In the single-target (control) condition, which allows a baseline assessment of the ability to detect T2, the T1 task does not have to be performed. Such a procedure enables T2 performance to be plotted as a function of the temporal interval between T1 and T2, as shown in Fig. 2.

prime T3 (presented after the blink) as indicated by better T3 performance when T2 and T3 were related, compared to when they were unrelated. This outcome suggests strongly that T2 reaches a high level of processing, even though it is not reportable. Similar conclusions were drawn earlier by both Marcel[11] and Lewis[12].

Using another approach, Luck *et al.*[7] (see also Ref. 8) examined event-related brain potentials (ERPs) elicited by T2. When T2 was unreportable during the AB interval, T2 elicited electrical potentials associated with early perceptual activity (N1 and P2 waveforms), as well as activity associated with meaning (N400 waveform). Interestingly, Vogel *et al.* showed that the waveform associated typically with the updating of working memory (P3, see Ref. 13) is absent during the AB. These dramatic findings support our belief that stimuli presented during the AB undergo a significant amount of processing, including that which facilitates meaning, but fall short of consciousness when this term is defined as meaning the ability to report successfully.

An interference theory of the AB

Shapiro and his colleagues[14–16] have postulated that interference in a short-term storage buffer is the source of the AB. Contributing to this interference are not only the two targets (T1 and T2) but also the items (masks) immediately following T1 and T2. Interference theory assumes that T1, T2 and their respective masks are processed to a varying degree and that these four items compete with each other, most likely during retrieval from this hypothetical storage buffer. Differential weighting of T1 over T2 (owing to T1's role as the first task) yields successful T1 report at the expense of T2.

As predicted by interference theory, the number of items competing in the storage buffer should affect the magnitude of the AB. Isaak *et al.*[15] find support for this prediction but only for stimulus items competing at a conceptual (meaning) rather than perceptual (feature) level. Isaak *et al.* used a variant of the RSVP method, as described in recent papers[17,18], where only the items critical to the AB effect are presented (T1, T2 and their respective masks) with varying T1–T2 intervals as required. In addition, the item just following T2's mask was presented to extend the range of the number of possible competitors. Isaak *et al.* presented combinations of letter and false-font (letter-like) stimuli randomly, yielding a minimum of two and a maximum of five 'real' letters per trial. The results, shown in Fig. 3, suggest that AB magnitude is a not just a function of the number of competing stimuli but, rather, the competitors must be from within the same conceptual category. Taylor and Hamm[19] showed the same outcome using digits and letters as the T2 task.

Alternative theoretical accounts

Duncan *et al.*[17] (see also Ref. 18) suggest that the AB represents the dwell time of attention. On the basis of neuropsychological evidence, Duncan[20] argues that during this dwell time attention acts to coalesce the properties of an object into a coherent perceptual representation. When viewed this way, the AB poses a serious challenge to a strict serial model of visual processing[21], where stimuli are said to be processed at rates of approximately 50 ms per item. However, Moore *et al.*[22] argue that the way in which serial models are interpreted in light of the outcome of visual search experiments may not be applicable to RSVP methodology.

Chun and Potter[23] propose a two-stage model to account for the AB. In stage 1, all stimuli are processed to a preliminary stage where features and even meaning are registered, but not at a level sufficient for report. In stage 2, stimuli are consolidated to a level required for a response. The second stage has been described as similar to the concept of working memory. According to their account, the AB occurs when stage 2 is unable to operate on T2 owing to this stage being occupied by T1.

Di Lollo and colleagues (pers. commun.) and Seiffert and Di Lollo[24] propose an account of the AB based on the role of T1 and T2 masks. Di Lollo and colleagues used both simultaneous masks (target and mask superimposed) and delayed masks (the mask follows the target) on both T1 and T2 in a fully crossed design. Their conclusion is that both kinds of T1 masking (simultaneous and delayed) act similarly when it comes to T2 performance: both cause a blink. However, when it comes to the masking of T2, they find that only masking by delay yields a blink. Such a conclusion argues that masking an attended object (T1) is different to masking an unattended (by virtue of the AB) object (T2). Furthermore, it is suggested that the lack of attention to T2 may allow object substitution of the T2 mask for T2. Enns and Di Lollo[25] argue for a similar account to explain masking in unattended visual locations. Evidence in support of such object substitution was reported by Isaak *et al.*[15] It is important to note that Isaak *et al.* find that, under certain circumstances, the T1 and its mask are as equally likely to be reported as T2 and the T2 mask. As a further note here, there is firm ground on which to argue that a mechanism such as object substitution underlies the phenomenon of induced change blindness reported recently by Rensink[26].

A fourth alternative account postulates that the AB may be related to the psychological refractory period (PRP)[27,28] and, indeed, PRP and AB methodologies bear a number of similarities. In PRP experiments, two targets are presented, separated by a variable interval. Participants are required to respond to the two targets as quickly as possible after each is presented. Two points differentiate these methods: firstly, PRP experiments require speeded responses to both targets (in AB experiments participants always respond, under no time pressure, after the stimulus stream has ended); and secondly, PRP targets are not masked (masks are required to produce an AB and although the second target in a PRP experiment may mask the first when it follows with a short [stimulus onset asynchrony (SOA), this is not a requirement to produce the typical PRP outcome]. The results of PRP experiments show that average response time to the second stimulus is slowed dramatically with short interstimulus intervals and that response times to both stimuli are positively correlated. AB-type experiments using PRP methods (speeded responses required after each target) have shown that T1 reaction time predicts T2 accuracy and that response selection factors (such as simple versus forced choice) can modulate AB magnitude. Given the similarity of the PRP to AB outcomes, Jolicoeur and Dell Acqua maintain that the AB may be viewed as a response selection-based phenomenon, such as PRP (pers. commun.).

A unified model

The following three tenets, common to these five theories, represent an attempt to unify these diverse positions. (1) As a result of the T1 mask, increased attention is required to enable T1 to reach a level of awareness sufficient for report. (2) As less attention is available for T2, by virtue of T1's demands, T2 cannot be consolidated into a durable storage sufficient for report. This leaves T2 vulnerable to decay and/or object substitution from a variety of stimulus sources, most notably from T2's mask but also, under appropriate conditions, by T1 as well as the T1 mask. This condition is attenuated gradually as T1's identity is resolved, following the approximately half second period which we refer to as the AB. In spite of the inability to report T2 with a high degree of accuracy during this interval, T2 is processed to a level of semantic awareness. (3) If the system is put under further constraint, for instance, by a rapid T1 response requirement, then response-selection factors will bear additionally on T2 accuracy.

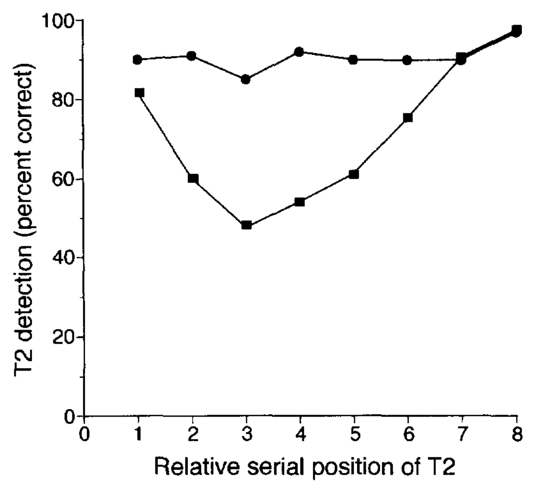


Fig. 2 Typical rapid serial visual presentation (RSVP) experimental results revealing the attentional blink (AB). Performance on T2 is plotted on the Y-axis and the serial position and time post-T1 are plotted on the X-axis. The AB is defined as the interval between 90 and 540 ms when performance in the dual-target condition (filled squares) is worse than in the single-target condition (filled circles).

Cross-modal AB

In this section, we examine recent evidence addressing the specific issue of cross-modal AB. The larger question here is what happens when we direct attention to targets from

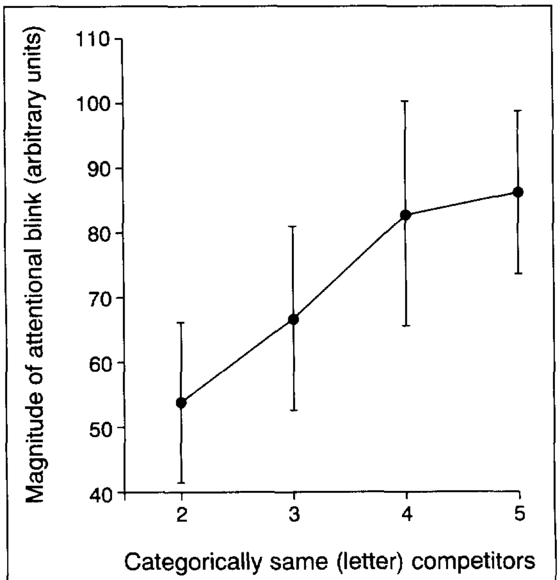


Fig. 3 The results of Isaak *et al.*[15], modelled after the procedure used originally by Duncan *et al.*[17] As can be seen, T2 performance, as reflected in the magnitude of the attentional blink, is a linearly increasing function of the number of letters appearing in the five-item rapid serial visual presentation (RSVP) display, suggesting that competition in (visual short-term memory) VSTM occurs only among conceptually similar stimuli.

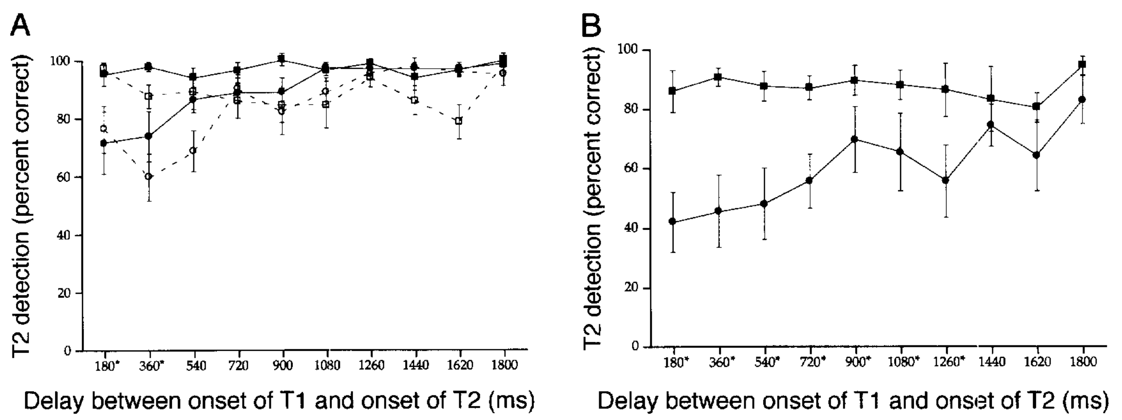


Fig. 4 The attentional blink paradigm as applied to a neuropsychological population. (A) Results of non-patients and patients without neglect (filled symbols and open symbols, respectively) tested under the single- and dual-target conditions (squares and circles, respectively) used in the study by Husain *et al.*[33] Control groups reveal a normal attentional blink. **(B)** Results of patients with visual neglect tested under the single- and dual-target (squares and circles, respectively) conditions used in the study by Husain *et al.*[33] A protracted attentional blink is observed in the dual-target condition lasting approximately three times the duration of controls shown in (A). (* Attentional blink.)

different modalities? The question is an interesting and timely one, as most current theories of the AB hold it to be a strictly visual phenomenon. A number of outcomes are possible: (1) there are a number of modality-specific attention mechanisms (for example, auditory and visual); (2) there is a single, non-modality-specific attentional mechanism, sometimes referred to as a supramodal mechanism (for example, neither auditory nor visual); (3) there is an interacting set of modality-specific attentional systems (for example, auditory affects visual but not the other way around); or (4) there is a single modality-specific attentional mechanism (for example, visual but not auditory).

The issue of cross-modal attention is being investigated using a variety of paradigms, such as exogenous cueing[29] and the attentional blink. Interest in cross-modal interactions is not new, with Posner and Boies[30] showing conditions under which both visual–auditory independence and dependence were observed. With regard to the AB, Potter and Chun have argued that the AB occurs only between visual targets and that other forms of attentional interference are independent of visual AB and in many cases reflect task switching demands[31]. On the other hand, Duncan *et al.*[32] present evidence in support of multiple, separate systems. Their data reveal the existence of only a visual–visual blink and an auditory–auditory blink, but no cross-modal effects. Arnell and Jolicoeur (unpublished) provide evidence for a single, supramodal mechanism, where cross-modal stimuli affect each other approximately as much as within-modality stimuli. The central assumption in Arnell and Jolicoeur's model is that encoding a stimulus into short-term memory requires central (supramodal) mechanisms that are capacity limited such that, usually, only one operation requiring those mechanisms can be performed at one time.

Conflicting evidence such as this requires a careful examination of the specific paradigms used; such an examination reveals significant differences. Potter and Chun, who found no evidence for either an auditory or cross-modal AB, used a relatively slow rate of stimulus presentation (135 ms SOA) and a single RSVP stream which makes their study similar to previous AB work. Duncan *et al.*, on the other hand, used concurrently presented auditory and visual streams. Arnell and Jolicoeur, also using concurrent streams, used faster rates of stimulus presentation (90 ms SOA), which they argue is particularly important to reveal an AB in the auditory modality. However, Arnell and Jolicoeur used a digit T1 and a letter T2, which Potter and her colleagues argue leaves open the possibility of task switching to account for their cross-modal effects. Although the weight of the evidence suggests that a cross-modal AB can exist, the conflicting results reported so far suggest that the circumstances promoting such effects remain to be fully specified.

Applications to neuropsychology

Recently, Husain and colleagues[33] have used the AB paradigm to examine the temporal dynamics of attention in individuals with visual neglect. Visual neglect is a common disorder following right hemisphere lesions caused by a stroke; it affects over 70% of patients with such lesions. These individuals are unaware of people or objects in their contralesional, or left, visual field whereas the ipsilesional or right visual field remains relatively unimpaired.

Prominent theories to explain visual neglect share the fundamental assumption of a spatial attentional deficit. For example, Posner *et al.*[34] have argued that neglect patients have a directionally specific impairment of disengaging attention from a stimulus in the ipsilesional field when they are required to shift attention to the contralesional field. Husain *et al.*[33] wanted to show that the impairment experienced by individuals with visual neglect might have a temporal attentional component to it: these patients might have an increased difficulty in disengaging from a stimulus, regardless of a need to move attention from one spatial location to another. The results of two control groups that showed no neglect (age-matched and lesion site-matched) are shown in Fig. 4A and reveal a normal AB profile. The performance of neglect patients is shown in Fig. 4B: their ability to detect T2 did not return to baseline until nearly 1.5 seconds, revealing an AB lasting approximately three times longer than controls. Such a finding suggests that individuals with visual neglect may suffer from a prolonged inability to disengage from T1, which in turn may serve to explain the nature of their visual deficits from a temporal, rather than a spatial, perspective.

Conclusions

The first AB report[3] was published five years ago. During these few short years the pursuit of the AB phenomenon has moved beyond a methodology that provides a window on the timecourse of object processing, towards a method for investigating what, arguably, is referred to as consciousness. If we define consciousness as the overt ability to report a specified stimulus, then we would have to state categorically that we are not consciously aware of the second target in an RSVP paradigm. If, on the other hand, we define consciousness as the ability to use the information contained in T2, then indeed we are very conscious of more than we know. The research presented here is consistent with the view that one of the roles of attention is to raise a stimulus to a conscious level. Conceptualizing attention in this way elevates it to a role of prominence in the processing of stimulus information. This is not a new conclusion but one certainly worth making again.

Acknowledgements

Some of the ideas presented in this paper have their roots in various informal discussions which took place during the 1997 Cognitive Science Association for Interdisciplinary Learning (CSAIL) conference in Hood River, OR, USA.

Outstanding questions

- Why is T2 accuracy often relatively good when T2 is presented within approximately 100 ms of T1 (see position 1 of Fig. 2)? Most theories of the AB explain such findings in terms of the adjacent T1 and T2 items being processed in a 'single temporal episode'. However, such explanations are *post-hoc* in nature and lack any suggestion of a plausible mechanism able to account for this outcome.
- The high T2 accuracy at very short T1–T2 intervals is not present in a significant number of AB studies. In these studies, T2 accuracy is worst at the very short T1–T2 intervals. Informal comparisons of studies suggest that the difference cannot be attributed to the difficulty of the T1 task, the nature of the T1 task (for example, detecting a fully specified T1 versus detecting the T1 feature and then reporting an orthogonal feature), the nature of the filler RSVP items or overall T2 accuracy. What is the key to understanding T2 accuracy at very short T1–T2 intervals?
- The AB is reduced or even eliminated when the T1+1 item and/or T2+1 items are removed. It appears that these target+1 items are important both conceptually and featurally, in terms of masking. Of interest is how featural masking exerts its affect on the AB. Does a T1 mask increase the amount of attention T1 requires to be recognized, whereas a T2+1 mask provides an item to be substituted for T2 (Refs 24, 25)? Do both the T1+1 and T2+1 items provide items to be substituted and/or confused with the targets[14–16]? Or, do the masks overwrite the iconic images of the targets, forcing the targets to be identified at the time that they are presented, thereby allowing their decay while they wait to gain access to the bottlenecked processor[23]?
- Although some of the critical factors accounting for the discrepant cross-modal AB findings have been elucidated, others have not. What are the critical factors that account for these remaining discrepancies?
- Recently, Joseph and colleagues[35] have demonstrated that an AB can be found for simple visual features such as orientation, thereby providing evidence that such features do require attention. This is in contrast to previous theories suggesting that such stimuli are preattentive, given that such simple visual features have been shown to 'pop out' of visual search displays, resulting in search times that are constant across a changing number of distractor items. Furthermore, the AB suggests that the cost of attending to an item is approximately 500 ms whereas the visual search results suggest a cost of approximately 50 ms. Can such findings be reconciled?
- Raymond and Sorensen have demonstrated that the AB is eliminated when induced motion leads the subject to interpret the T2 target as being the same object as the T1 target (pers. commun.). These authors suggested that the absence of AB in their experiment resulted from subjects not needing to open an object file for T2, as is usually required, and that the ability to create new object files may be reduced during the AB. Can the notion of 'object files', taken from the visual search literature, contribute to our understanding of the AB?
- As mentioned above, some researchers have suggested that the process of object substitution is vital to the production of the AB. Is it possible that the same process of object substitution results in the phenomenon of induced change blindness reported recently by Rensink[26]?

References

1 Lawrence, D.H. (1971) Two studies of visual search for word targets with controlled rates of presentation *Percept. Psychophys.* 10, 85–89

2 Broadbent, D.E. and Broadbent, M.H.P. (1987) From detection to identification: Response to multiple targets in rapid serial visual presentation *Percept. Psychophys.* 42, 105–113

3 Raymond, J.E., Shapiro, K.L. and Arnell, K.M. (1992) Temporary suppression of visual processing in an RSVP task: An attentional blink? *J. Exp. Psychol. Hum. Percept. Perform.* 18, 849–860

4 Weichselgartner, E. and Sperling, G. (1987) Dynamics of automatic and controlled visual attention *Science* 238, 778–780

5 Reeves, A. and Sperling, G. (1986) Attention gating in short term visual memory *Psychol. Rev.* 93, 180–206

6 Shapiro, K.L. *et al.* (1997) Priming from the attentional blink: A failure to extract visual tokens but not visual types *Psychol. Sci.* 8, 95–100

7 Luck, S.J., Vogel, E.K. and Shapiro, K.L. (1996) Word meanings can be accessed but not reported during the attentional blink *Nature* 382, 616–618

8 Vogel, E.K., Luck, S.J. and Shapiro, K.L. Electrophysiological evidence for a post-perceptual locus of suppression during the attentional blink *J. Exp. Psychol. Hum. Percept. Perform.* (in press)

9 Maki, W.S., Frigen, K. and Paulson, K. (1997) Associative priming by targets and distractors during rapid serial visual presentation: Does word meaning survive the attentional blink? *J. Exp. Psychol. Hum. Percept. Perform.* 23, 1014–1034

10 Shapiro, K.L., Caldwell, J.I. and Sorensen, R.E. (1997) Personal names and the attentional blink: The 'cocktail party' effect revisited *J. Exp. Psychol. Hum. Percept. Perform.* 23, 504–514

11 Marcel, A.J. (1983) Conscious and unconscious perception: Experiments on visual masking and word recognition *Cognit. Psychol.* 15, 197–237

12 Lewis, J. (1970) Semantic processing of unattended messages during dichotic listening *J. Exp. Psychol.* 85, 225–228

13 Donchin, E. (1981) Surprise!... Surprise? *Psychophysiology* 18, 493–513

14 Shapiro, K.L., Raymond, J.E. and Arnell, K.M. (1994) Attention to visual pattern information produces the attentional blink in RSVP *J. Exp. Psychol. Hum. Percept. Perform.* 20, 357–371

15 Isaak, M.I., Shapiro, K.L. and Martin, J. The attentional blink reflects retrieval competition among multiple RSVP items: Tests of the interference model *J. Exp. Psychol. Hum. Percept. Perform.* (in press)

16 Raymond, J.E., Shapiro, K.L. and Arnell, K.M. (1995) Similarity determines the attentional blink *J. Exp. Psychol. Hum. Percept. Perform.* 21, 653–662

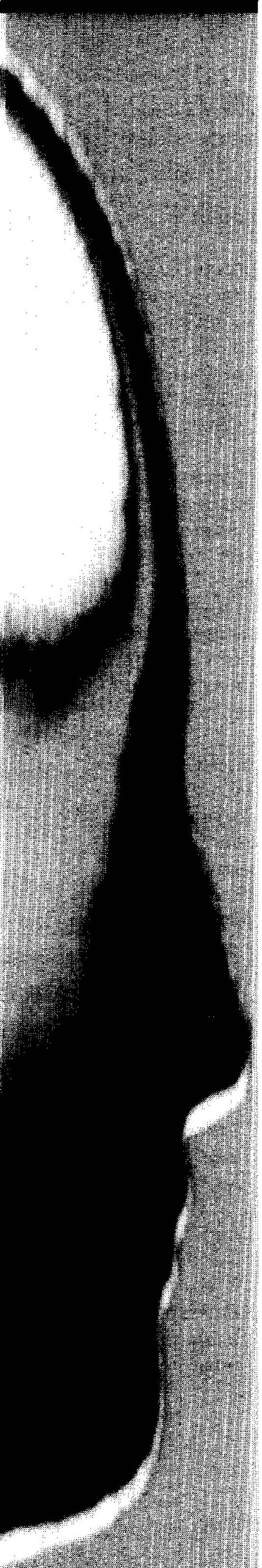
17 Duncan, J., Ward, R. and Shapiro, K.L. (1994) Direct measurement of attentional dwell time in human vision *Nature* 369, 313–315

18 Ward, R., Duncan, J. and Shapiro, K.L. (1996) The slow time-course of visual attention *Cognit. Psychol.* 30, 79–109

19 Taylor, T.L. and Hamm, J. (1997) Category effects in temporal visual search *Can. J. Exp. Psychol.* 51, 36–46

20 Duncan, J. (1996) Co-ordinated brain systems in selective perception and action, in *Attention and Performance XVI* (Inui, T. and McClelland, J.L., eds), pp. 549–578, MIT Press

21 Treisman, A.M. and Gelade, G. (1980) A feature-integration theory of attention *Cognit. Psychol.* 12, 97–136
22 Moore, C.M. *et al.* (1996) Are attentional dwell times inconsistent with serial visual search? *Psychonomic Bull. Rev.* 3, 360–365
23 Chun, M.M. and Potter, M.C. (1995) A two-stage model for multiple target detection in rapid serial visual presentation *J. Exp. Psychol. Hum. Percept. Perform.* 21, 109–127
24 Seiffert, A.E. and Di Lollo, V. (1997) Low-level masking in the attentional blink *J. Exp. Psychol. Hum. Percept. Perform.* 23, 1061–1073
25 Enns, J.T. and Di Lollo, V. (1997) Object substitution: A new form of masking in unattended visual locations *Psychol. Sci.* 8, 135–139
26 Rensink, R.A., O'Regan, J.K. and Clark, J.J. (1997) To see or not to see: The need for attention to perceive changes in scenes *Psychol. Sci.* 8, 368–373
27 Pashler, H. (1994) Dual-task interference in simple tasks: Data and theory *Psychol. Bull.* 2, 220–244
28 Welford, A.T. (1952) The 'psychological refractory period' and the timing of high-speed performance – A review and theory *Br. J. Psychol.* 43, 2–19
29 Spence, C. and Driver, J. (1997) Audiovisual links n exogenous covert spatial orienting *Percept. Psychophys.* 59, 1–22
30 Posner, M.I. and Boies, S.J. (1971) Components of attention *Psychol. Rev.* 78, 391–408
31 Potter, M.C. and Chun, M.M. Two attentional deficits in serial target search: The visual attentional blink and an amodal task-switch deficit *J. Exp. Psychol. Learn. Mem. Cog.* (in press)
32 Duncan, J., Martens, S. and Ward, R. (1997) Restricted attentional capacity within but not between sensory modalities *Nature* 369, 313–315
33 Husain, M. *et al.* (1997) Temporal dynamics of visual attention reveal a non-spatial abnormality in spatial neglect *Nature* 385, 154–156
34 Posner, M.I. *et al.* (1987) How do the parietal lobes direct covert attention? *Neuropsychologia* 25, 135–145
35 Joseph, J.S., Chun, M.M. and Nakayama, K. (1997) Attentional requirements in a 'preattentive' feature search task *Nature* 387, 805–808



Computational theories of object recognition

Shimon Edelman

This paper examines four current theoretical approaches to the representation and recognition of visual objects: structural descriptions, geometric constraints, multidimensional feature spaces and shape-space approximation. The strengths and weaknesses of the four theories are considered, with a special focus on their approach to categorization – a computationally challenging task which is not widely addressed in computer vision, where the stress is rather on the generalization of recognition across changes of viewpoint.

S. Edelman is at the School of Cognitive and Computing Sciences, University of Sussex, Falmer, Brighton, UK BN1 9QH.

tel: +44 1273 678659
fax: +44 1273 671320
e-mail: shimone@cogs.susx.ac.uk

The study of visual object recognition has seen such rapid development recently that its comprehensive survey would not fit within the confines of a journal paper. In this short review I concentrate on some aspects of what Marr[1] termed the 'computational theory' of object representation. Recognition algorithms stemming from the different computational formulations of the problem of representation are also mentioned. Very little space is devoted to implementational issues, and none at all to the evaluation of various theories as models of human performance or as explanations of the functional neurobiology of object recognition in primates (see Refs 2–5 and the forthcoming special issues of *Vision Research* and *Cognition*). Traditionally, in cognitive science, debates concerning theories of object representation centre on computational problems stemming from the effect of viewpoint on the appearance of objects[6–12]. The emergence of powerful formal methods for overcoming the effect of viewpoint[7,13,14] and the recent successes of surprisingly simple empirical approaches to recognition[15,16] are likely to shift the focus of theoretical discussion to other topics. Indeed, my chief aim here is to bring to the foreground a class of computational problems that differ from those related to viewpoint dependence, yet confront any recognition system. These problems arise from the need to 'categorize', or make sense of, novel objects.

Perception, recognition and categorization

The spectrum of problems arising in connection with object recognition is best understood in terms of two basic distinctions. The first of these has to do with the perception of the shape of an observed object on the one hand, and the recognition of objects on the basis of their shapes on the other hand. A classical observation of this distinction was made by Wittgenstein[17], who discussed at length the difference between seeing a shape and seeing it 'as something'. Unlike merely perceiving a shape (a problem not addressed in the present review), recognizing it as something involves memory, that is, representations of shapes seen in the past. The form of these representations is constrained by the various factors, such as orientation and illumination, that affect the appearance of objects. Because of the effects of orientation, for instance, simply storing a particular snapshot of an object for future reference would not do: another view of the same object may turn out to be less similar to the stored view than to a view of a different object, leading to an erroneous recognition. As noted above, contemporary theoretical treatments

 1364-6613/97/$17.00 PII: S1364-6613(97)01090-5