Early Top-Down Influences in Control of Attention: Evidence from the Attentional Blink

Frederic Hilkenmeier, Jan Tünnermann, and Ingrid Scharlau

Paderborn University, Warburger Str. 100, 33098 Paderborn, Germany
frederic.hilkenmeier@uni-paderborn.de

Abstract. The relevance of top-down information in the deployment of attention has more and more been emphasized in cognitive psychology. We present recent findings about the dynamic of these processes and also demonstrate that task relevance can be adjusted rapidly by incoming bottom-up information. This adjustment substantially increases performance in a subsequent task. Implications for artificial visual models are discussed.

Key words: visuo-spatial attention, top-down control, task relevance, artificial visual attention, attentional blink

1 Introduction

According to widespread assumptions in cognitive psychology and modeling research, attention facilitates processing of attended compared to unattended information. This is achieved by selecting relevant or conspicuous information or disregarding information which is not relevant to the current task or visually unobtrusive. Basically, attended information is processed in more detail [1] or quicker [2] than unattended information. In contrast to unattended information it can be part of higher level object representations [3], or be perceived consciously [4]. Although the general finding - a boost of attended information - is undisputed, the control mechanisms are a matter of debate. How far, for instance, is attention controlled by visual saliency of a stimulus, and how much can be done by current intentions [5] [6]? Here, we want to draw attention towards two current lines of thinking. First, attending to a stimulus may come by a cost that is overlooking information in the close temporal vicinity of an attended target (attentional blink [7]. Second, the role of top-down task relevance in the control of attention cannot be stressed too much [8] [9], at least after the initial transient bottom-up saliency [10]. In an attentional blink task, observers watch simple dynamic scenes consisting of a fast stream of single items in the same spatial location, as illustrated in Figure 1. These items are either distractors (e.g. letters) or targets (e.g. digits). In this setup, observers typically have no difficulties reporting the first target digit ($T1$), but fail in identifying the second target ($T2$) if it is presented between 200 to 500 ms after $T1$. It is as if attention, analogous to the lid closure of an eye blink, briefly switched off before new information can be processed [11]. The precise adaptive significance behind this attentional deficit is still unknown. Interestingly, attention seems to consist of two mechanisms: a boost (facilitation) after a target has been detected, and a bounce (inhibition) after a distractor has been detected. $T2$ suffers from the attentional inhibition, which was originally directed to the distractor preceding $T2$. The approach to define attention as two-folded is quite novel [9], but is nevertheless already impressive in explaining a number of empirical results from different spatial and temporal phenomena. More important in the present context is that this attentional allocation depends upon the current task, not on bottom-up input like intensity or color. Of course, the cognitive system requires this bottom-up input to decide whether a stimulus matches the task set or not, but the deployment of attention is modulated by task relevance: stimuli that match the criterion get boosted, whereas stimuli that do not match the task set (or even match a distractor set; see [12] get bounced. This importance of task relevance was impressively shown by Nieuwenstein [13]: A distractor preceding $T2$ that shares a feature with $T2$ can significantly reduce the attentional blink deficit. For example, a red distractor letter preceding a green digit $T2$ is an effective cue when the task is to look for red and green digits, but the same red cue is ineffective when the task is to look for only green digits. These recent findings in cognitive psychology emphasize task relevance as key element for attentional deployment: depending on the current task set, the exact same stimuli can either be inhibited or facilitated. In contrast, classic computational models [14] [15] understand spatial attention as a stimulus-driven, bottom-up process of feature-integration [16]. The input scene is processed in maps of simple features. In the Itti and Koch model[14], these are color, intensity, and orientation, but others may be appropriate, such as symmetry, size, and eccentricity in the model of Aziz and Mertsching [15]. Within each map, stimuli with a certain singularity compared to their neighbors are more salient and therefore receive higher values. These single feature maps are combined into one saliency map which can be used to determine the most conspicuous location in the scene and guide attention. So far no explicit mechanisms integrate top-down influences, such as task relevance. The mentioned models can be viewed as hardwired for the task "find the most conspicuous feature" with no further knowledge and influences. A weighted combination of feature maps can bias the system in favor for certain features, tweaking the model to fulfill more specific tasks like: "Find a conspicuous feature with a high color saliency ignoring orientation". Navalpakkam and Itti [17] suggest fine-grained weighting, allowing biases not only in favor for the feature in general, but also for specific intervals. Again, the model can be configured to perform narrower tasks like: "Find a dot of intermediate intensity." Setting up weights, along with further model parameters, such as the constants depicted in table 1 in [15], provides a priori task configurability. For many applications in computer vision the described models may be sufficient. However, from a biological point of view they do not fully depict the early and parallel influences of top-down knowledge. Here, we want to demonstrate that a specific stimulus which has relevant information for a subsequent task can rapidly be processed and used for top-down allocation of attention within the task. This means that the top-down allocation can be adjusted dynamically within virtually no time, which is clearly at odds with a rather static view of saliency maps.

2 Experiment

In the present attentional blink experiment, we integrate information about the temporal position of target $T2$ into the identity of target $T1$. The targets are digits, with $T1$ being a valid cue informing about the target onset asynchrony (i.e. the specific temporal onset of $T2$). There are two conditions: In the baseline condition the participants are unaware of the fact that the cue informed about the target onset asynchrony. In the top-down condition, the instructions emphasize the relationship between $T1$ identity and temporal position of $T2$. A diminished attentional blink in the top-down condition would support our idea that the identity of a leading target stimulus is rapidly available for top-down usage in attention. This is found in other paradigms and with other dependent measures like temporal order judgments [18], event-related brain potentials, reaction times [6], or MEG data [19] as well.

2.1 Apparatus, Stimuli, and Procedure

Participants sat in a dimly lit room with the center of a 19" CRT monitor at eye level. Viewing distance was 50 cm, set by a chin rest. The resolution was 800 x 600 pixels at 100 Hz. The experimental program was written in MATLAB 7.5.0 and made use of the PsychToolbox [20]. Stimuli were displayed as black symbols on a medium grey background. The distractors were letters from the Roman alphabet; the targets were the digits 1-9. Stimuli subtended about 1 in visual angle. Each item was presented for 40 ms and separated from the next by an interstimulus interval of 10 ms, resulting in a presentation rate of 20 items/sec. Stimuli were presented at the center of the screen (Figure 1). The forty students that participated in this experiment and were paid for their participation were split into two groups. In both groups, the first target digit was a 100% valid cue about the lag $T2$ was presented in. This means when $T1$ was a "1", $T2$ immediately followed $T1$ (lag1). When $T1$ was a "2", $T2$ was presented at lag2, with one intervening distractor between the targets, and so on. The identity of the second target was selected randomly. The instruction for the 20 observers in the top-down group emphasized these facts, whereas the 20 observers in the baseline group were told that the two target digits were always random. The lags 1,2,3,5 and 6 were included in this experiment, that is, the maximal target onset asynchrony was 300 ms. Each lag was repeated 20 times. A fixation cross was presented at the center of the screen for 200 ms followed by a rapid serial visual presentation (RSVP) stream consisting of 26 items. The first target was the 11th to 15th character in a stream. The second target was followed by 4 to 14 distractors. The targets within a trial were always different. After each trial, the observers were asked to identify the two targets in the order they appeared in by

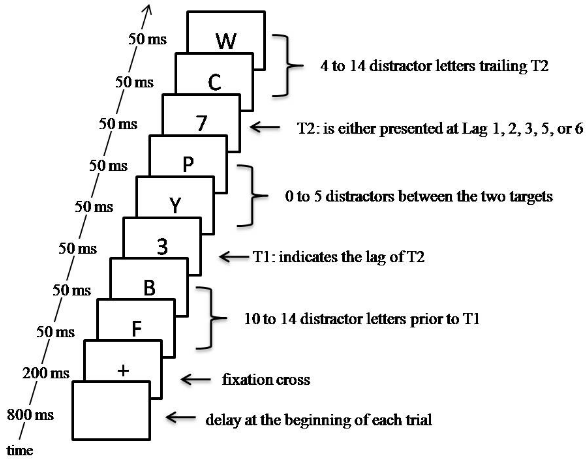

Fig. 1. Example of stimulus sequences. Here, $T2$ follows $T1$ at lag3 (after 150ms).

pressing the corresponding keys on the keypad. In case they had not recognized one or both targets, they were encouraged to guess. Observers were allowed to rest whenever they wanted.

3 Results and Discussion

Figure 2 shows the identification rate of $T2$ given $T1$ identified correctly ($T2|T1$), separately for the top-down and baseline conditions. The attentional blink is attenuated, but not abolished, when participants are instructed to use the identity information of the first target as a cue for the temporal position of the second target. This means that observers have fast access to the identity of the first target and were also able to use this information for a virtually instant redeployment of attention. As can be seen from Figure 2, the first significant difference between the top-down and the baseline condition is at lag 2, that is 100 ms after $T1$ onset. We therefore can conclude that the information about $T1$ is available to top-down usage within 100 ms.

3.1 Statistics

In statistical terms the attenuated attentional blink is shown by a significant main effect for the within-subjects factor of lag ($F[4, 190] = 23.97, p = 0$) and also a significant main-effect for the between-subjects factor condition (top-down vs. baseline; $F[1, 190] = 8.64, p = 0.004$) in a two way analysis of variance. The interaction was not significant ($F[4, 190] = 0.39, p = 0.81$). Additional independent one-tailed two-sample t-tests between the top-down and the baseline condition revealed a significantly better $T2|T1$ performance at the lags 2 and 5 ($t[38] = 1.74$, $p = 0.04$ and $t[38] = 1.7$, $p = 0.4$, respectively).

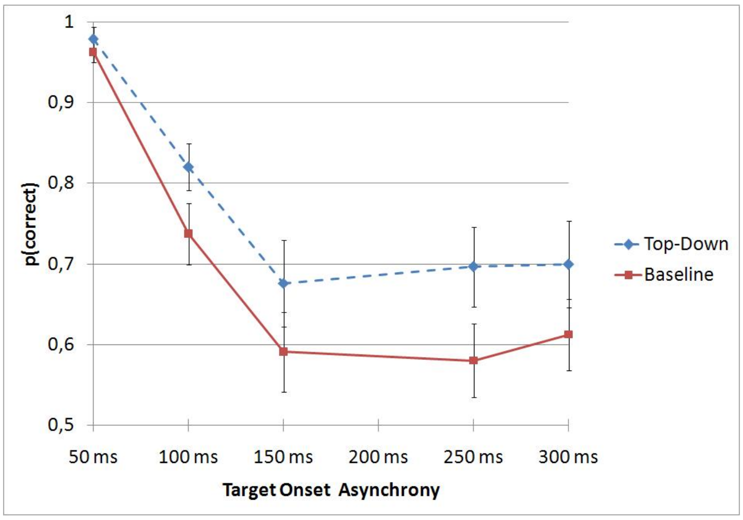

Fig. 2. Conditional accuracy of $T2|T1$ as a function of lag in steps of 50 ms. Order reversals are counted as correct.

4 General Discussion

To sum up: Intentions play a vital role in the control of attention. Top-down information kicks in very early in information processing. This is a central finding in actual psychological discussion and has been shown in other attentional paradigms like spatial cueing as well: A feature singleton only captures attention when its feature matches a currently active task set, but not when it is irrelevant for the task [5]. These findings clearly indicate that attention is modulated solely [6], or at least primarily [5] [18], via top-down information about the current task. But task demands not only influence which stimuli get preferred processing. In the present experiment we demonstrated that the task can be adjusted dynamically. This top-down adjustment is due to the characteristics of a stimulus that is just being processed, indicating that the brain is capable of rapid evaluation of a stimulus' behavioral relevance. These early attentional effects are not considered in the computational models discussed so far and are, at least in our view, difficult to implement.

Hamker's model [21] [22] takes a somewhat different approach: To account for rapid attentional effects, he proposes a computational attention model which utilizes feature feedback and spatially organized reentry to implement task relevance in a way that matches biological observations. Bottom-up information from feature maps meet a target template which is held in a functional block representing working memory. This initiates a dynamic recognition process and enhances features that match the target template and filters out information that is irrelevant when fed back to the original feature maps. This happens for all feature dimensions (such as orientation and color). The activation in feature maps is also projected to a functional block, which simulates a brain areal that programs eye movements in human vision. Here a competition between the active regions takes place, until the system stabilizes with high activation in one region and low everywhere else. This information reenters the feature map, now enhancing specific locations in addition to the feature specific enhancements mentioned before. This step-by-step description might lead to misinterpretation of the model as a sequential system, which it is not. All components act in parallel and continuously, implicitly emerging attention.

Hamker's intention to investigate in biological plausible top-down interaction, rather than developing a usable model for computer vision, results in a minimal model to run on simplified inputs. Because of that, direct application in computer vision is not possible. Nevertheless, Hamker and colleagues [23] [19] have successfully tested this model in experimental work.

In conclusion, vision is an active, dynamic and constructive process, as demonstrated by a variety of findings presented throughout this article. If stimuli provide information that is relevant for the current task, this information can rapidly be used to adjust the task and substantially increase performance. In our view, this shows that investigating the nature of top-down influences is valuable for a biologically inspired model of artificial visual attention.

Acknowledgments. Supported by DFG grants Ne 366/7-2 to Odmar Neumann and Scha 1515/1-1 to Ingrid Scharlau.

References

1. Broadbent, D.E.: Perception and Communication. Pergamon, London (1958)
2. Posner, M.I.: Orienting of attention. The Quarterly Journal of Experimental Psychology **32**(1) (1980) 3–25
3. Kahneman, D., Treisman, A., Gibbs, B.J.: The reviewing of object files: Object-specific integration of information. Cognitive Psychology **24**(2) (April 1992) 175–219
4. Simons, D.J., Chabris, C.F.: Gorillas in our midst: sustained inattentional blindness for dynamic events. Perception **28**(9) (1999) 1059–1074
5. Folk, C.L., Remington, R.W., Johnston, J.C.: Involuntary covert orienting is contingent on attentional control settings. Journal of experimental psychology. Human perception and performance **18**(4) (November 1992) 1030–1044
6. Ansorge, U., Kiss, M., Eimer, M.: Goals-driven attentional capture by invisible colours: Evidence from event-related potentials. Psychonomic Bulletin & Review. (in press)
7. Chun, M.M., Potter, M.C.: A two-stage model for multiple target detection in rapid serial visual presentation. Journal of Experimental Psychology: Human Perception and Performance **21**(1) (February 1995) 109–127
8. Ansorge, U., Neumann, O.: Intentions determine the effect of invisible metacontrast-masked primes: Evidence for top-down contingencies in a peripheral cuing task. Journal of Experimental Psychology: Human Perception and Performance **31** (2005) 762–777
9. Olivers, C.N., Meeter, M.: A boost and bounce theory of temporal attention. Psychological review **115**(4) (October 2008) 836–863

10. Van Zoest, W., Donk, M.: Goal-driven modulation as a function of time in saccadic target selection. The Quarterly Journal of Experimental Psychology **61**(10) (2008) 1553–1572
11. Raymond, J.E., Shapiro, K.L., Arnell, K.M.: Temporary suppression of visual processing in an rsvp task: an attentional blink? . Journal of Experimental Psychology: Human Perception and Performance **18**(3) (August 1992) 849–860
12. Olivers, C.N.L., Watson, D.G.: Input control processes in rapid serial visual presentation: Target selection and distractor inhibition. Journal of Experimental Psychology: Human Perception and Performance **32** (2006) 1083–1092
13. Nieuwenstein, M.R.: Top-down controlled, delayed selection in the attentional blink. Journal of Experimental Psychology: Human Perception and Performance **32** (2006) 973–985
14. Itti, L., Koch, C.: Computational modelling of visual attention. Nature Reviews Neuroscience **2**(3) (Mar 2001) 194–203
15. Aziz, M.Z., Mertsching, B.: Fast and robust generation of feature maps for region-based visual attention. Image Processing, IEEE Transactions on **17**(5) (2008) 633–644
16. Treisman, A.M., Gelade, G.: A feature-integration theory of attention. Cognitive Psychology **12**(1) (January 1980) 97–136
17. Navalpakkam, V., Itti, L.: Top-down attention selection is fine-grained. Journal of Vision **6**(11) (Oct 2006) 1180–1193
18. Scharlau, I., Ansorge, U.: Direct parameter specification of an attention shift: evidence from perceptual latency priming. Vision Res **43**(12) (June 2003) 1351–1363
19. Bröckelmann, A.K., Junghöfer, M., Scharlau, I., Hamker, F.H.: Reentrant processing from attentional task sets: Converging support from magnetoencephalography and computational modeling. Submitted to Proceedings of the National Academy of Sciences of the United States of America
20. Brainard, D.H.: The psychophysics toolbox. Spatial vision **10**(4) (1997) 433–436
21. Hamker, F.: A computational model of visual stability and change detection during eye movements in real-world scenes. Visual Cognition (August 2005) 1161–1176
22. Hamker, F.H.: Modeling feature-based attention as an active top-down inference process. Bio Systems **86**(1-3) (2006) 91–99
23. Hamker, F.H.: A dynamic model of how feature cues guide spatial attention. Vision research **44**(5) (March 2004) 501–521