PERSPECTIVE

How Top-Down is Visual Perception?

featuring new data (VSS Poster):

"Attentional Cycles in Detecting Simple Events within Complex Displays"

Sunday PM Poster #36.301, VSS 2014

Thomas Sanocki, U. of South Florida

Summary

Everyday scene perception has strong top-down influences, but we need to look for them!
A perspective is developed here and example experiments introduced (poster 36.301).

Big questions about human perception remain quite open, including that of how top-down perception of the everyday world is. The concepts of top-down and bottom-up processing have been fundamental to perceptual research throughout its history. Yet, the overall influence of one or the other in everyday perception is unknown, as will be explained.

Top-down influences are more than a simple weight; top-down mechanisms interact with incoming data and have the potential for profound and lasting influences. For example, a top-down schema determines what types of information are most likely to be picked up and interpreted, and the information in turn influences learning and subsequent behavior (e.g., Neisser, 1976).

The present work begins with the idea that observers deal with a complex world through perceptual schemata that are modified over time. Here, a schema is an attentional control process, or attentional set. A set is necessary for the most efficient levels of identification and action to be realized. The set guides attention and processing toward a goal. When a specific set is active, events consistent with it are perceived efficiently, while inconsistent events are less likely to be perceived. Changing the set requires central attention, with costs over time. The costs can be large, and brief or long lasting, depending on attention.

The efficiency of data-driven processing

Data-driven processing is effective and important. However, strict bottom-up models are incomplete. Bottom-up models claim that highly efficient perception of familiar stimuli occurs independent of set. In a typical bottom-up model, stimulus features activate a pre-existing feedforward network for mapping features to response. This neural network operates independent of contextual factors, with the exception of temporary priming or bias effects. The model correctly predicts that identification is efficient for letters and words (e.g., Schneider & Shiffrin, 1977, Shapiro, Driver, Ward, & Sorensen, 1997), and receives support from the recent discovery that novel scenes are put into categories with great efficiency (e.g., Greene & Oliva, 2009, Kirchner & Thorpe, 2006).

Another bottom-up effect is that identification can occur outside of an observer's intentions and awareness (e.g., Mack, Pappas, Silverman, & Gay, 2002, Schneider & Shiffrin, 1977, Shapiro et al., 1997). Indeed, stimuli can capture observer attention away from other goals or sets (e.g., Franconeri & Simons, 2003, Yantis & Jonides, 1984). Such results support feedforward models of scene perception (e.g., Greene & Oliva, 2009, Henderson & Hollingworth, 2003, Kirchner & Thorpe, 2006, Schneider & Shiffrin, 1977).

The body of existing evidence is limited, however. Specifically, the challenges of the complexity of everyday scenes and task behavior have not been appreciated nor measured in most existing experiments. The interpretation of the evidence changes drastically in the perspective of everyday scene perception.

Everyday Scenes versus Simplicity-Biases

Everyday scene perception refers to the stimuli and processes that are typical of everyday activity within meaningful locations (natural scenes). It is an ideal domain for models of brain activity. In natural scenes, *stimulus complexity* is very high, not only at image levels, but also at the level of possible interpretations (e.g., Tsotsos, 1991). *Task complexity* is also very high; a wide range of tasks is possible within a scene, varying in content, type, and spatial and temporal scale.

The concept of selective attention was identified early in psychological science, motivated by complexity in everyday perception: How do observers select which of multiple possible stimulus streams they will interpret (James, 1890)? Internally driven (top-down) processes are critical for selection.

In contrast to the high levels of complexity of natural scenes and tasks, most of the experiments evaluating bottom-up and top-down processing have been conducted with simpler displays and tasks. Simplification is necessary for science; simple paradigms are easier to create and control. Consequently, few paradigms address the complexity of natural scenes. Varying display size to increase complexity is not enough to approach natural scene complexity. This has created a strong bias in the empirical literature — much more is known about simple display conditions than complex ones. Yet, complexity may be *the* primary reason for attention, control processes, and top-down mechanisms (e.g., James, 1890).

Although the bulk of the literature is strongly biased, researchers are beginning to explore complexity, and their results qualify the idea of bottom-up efficiency. Scene categorization is no longer highly efficient when the scene layout varies from trial to trial, with non-relevant objects present in the foreground (Walker, Stafford, & Davis, 2008). Letters are identified efficiently, but not in a "cost-free" manner; the costs of preparing attentional set can be observed shortly before identification (e.g., Paap & Ogden, 1981). Perceiving familiar stimuli is likely to be obligatory, running to completion once initiated, but also capacity limited (e.g., Lavie, Hirst, De Fockert, & Viding, 2004). The capture of attention is greatly reduced when displays start to become complex (Cosman & Vecera, 2009).

Experimental biases for simplicity go beyond stimulus factors. Researchers usually measure behavior during test periods that are preceded by practice. Yet, attention is set during initial practice periods, and initial portions of the experiment can be very interesting (e.g., Mack & Rock, 1998). In the Mack and Rock experiments, the unexpected stimulus is presented on the third trial along with a repetition of the expected (and difficult) task. This research indicates that mundane unexpected stimuli, such as a square, can be missed a substantial portion of the time (Mack and Rock, 1998). However, highly meaningful stimuli, such as a face or the observer's name, are often consciously detected, even though unprepared for. This result supports the idea of data-driven processing efficiency, and the obligatory processing of highly familiar stimuli to conscious levels. However, there also is a top-down effect involving the familiar stimuli — they are not processed as efficiently as if they were expected. For example, 88% of Mack et al.'s observers noticed their name despite inattention, whereas noticing should be near 100% if one's name is expected. This difference could reflect the full preparation of an attentional set for encoding a name.

The brain often responds to complexity by re-using mechanisms over time. However, most cognitive research focusses on single trials, missing much of the savings from re-use. A continuous stimulus stream is presented in attentional blink experiments (although spatial complexity is reduced by using a single location). Attentional blink experiments indicate that there are severe limits related to the temporal processing of stimuli; in particular, identifying a target stimulus creates high costs for subsequent targets that do not fit in the same set as the first target (e.g., Dux & Marois, 2009).

In summary, results supporting efficient bottom-up processing are greatly moderated by complexity, although the full extent of moderation is not known because stimulus and task variables that approach the complexity of natural scenes have barely been examined. Evidence suggests that the most efficient levels of processing are achievable only after a set has been instantiated.

Events in Space and Time

The complexity of natural scenes increases exponentially when the dimension of time is added. Most cognitive research in scene perception has been conducted with static stimuli, however. Stimuli that change in time pose new burdens on the processing system. One way to vary space and time is to create simple events that exist over seconds. Their temporal nature requires top-down guidance of attention across time.

In the present experiments, the events were simple, similar, and predictable; all had the same time scale, trajectory, and response. To create moderate complexity, multiple event tokens were presented, with staggered onset times. The events formed a stimulus stream that required continuous monitoring and responding.

Each trial included 47 distractor events and 7 target events. Each event was a variation of a walking stick figure (Figure 1). One target was distinguished by a hand clap (motion event), and the other by a color change (color brightening event). Distractors changed less (arms swung and color changed slightly). Observers were instructed to watch for both target events. Set was manipulated by varying the frequency of the target types. When one target type dominated, an attentional set for that event type developed. Consistent with models of visual search, attentional set can be thought of as a search set, for guiding attention to a candidate target and comparing it to a dynamic search template.

The goal of the experiments was to examine the activity of attentional set over time, in response to changes in the target events. Conditions of full attention to new targets (task changes) as well as reduced attention were examined, to test the idea that change of set depends on attention.

Preview of Results

The experiments confirm that a set is induced when one target type predominates, and that the effect is large and significant. When the other target type is introduced, a change of set is required and performance is markedly reduced. When there is full attention to the new targets (Experiment 1), the instantiation of the new set occurs quickly (in one of the extended trials), but is costly — a 34% reduction in identification. When new targets are introduced away from attention (Experiment 2), the cost is larger (42%) and the set is instantiated slowly; deficits continue thru 12 extended trials. When two sets are required (new targets introduced with old targets; Experiment 3), there large costs extending thru 12 extended trials.

This paper presented as part of VSS 2014 Poster

#36.301 (Inattentional blindness):"How Top-Down is Perception? Attentional Cycles in Detecting Simple Events in Complex Displays"

Poster available at: http://shell.cas.usf.edu/~sanocki/publicationspage.html

References

Cosman, J. D., & Vecera, S. P. (2009). Perceptual load modulates attentional capture by abrupt onsets. Psychonomic Bulletin & Review, 16, 404-410.

Dux, P. E., & Marois, R. (2009). The attentional blink: A review of data and theory. *Attention, Perception, & Psychophysics, 71*, 1683–1700.

Franconeri, S. L. & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics, 65*(7), 999-1010.

Henderson, J. M., & Hollingworth, A. (2003). Eye movements, visual memory, and scene representation. In M. A. Peterson and G. Rhodes (Eds.), *Analytic and holistic processes in the perception of faces, objects, and scenes (pp. 356-383).* New York: Oxford University Press.

James, W (1890). *The principles of Psychology* (Chapter XI: Attention).

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman, R. Davies, & J. Beatty (Eds.), *Varieties of attention* (pp. 29–61). New York: Academic Press.

Kirchner, H., & Thorpe, S. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, *46*(11), 1762-1776.

Lavie, N., Hirst, A., De Fockert, J. W. & Viding, E. (2004) Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General, 133*, 339-354.

Mack, A. & Rock, I. (1998). *Inattentional Blindness*. Cambridge, MA: MIT Press.

Mack, A., Pappas, Z., Silverman, M., & Gay, R. (2002). What we see: Inattention and the capture of attention by meaning. *Consciousness and Cognition*, *11*(4), 488-506. doi:10.1016/S1053-8100(02)00028-4.

Neisser, U. (1976). *Cognition and reality*. Freeman and Co.

Paap, K. R., & Ogden, W. C. (1981). Letter encoding is an obligatory but not capacity-demanding operation. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 518-527.

Schneider, W. & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84* (1), 1-66.

Shapiro, K and Driver, J and Ward, R and Sorenson, RE (1997) Priming from the attentional blink: A failure to extract visual tokens but not visual types. *Psychological Science, 8*, 95 - 100.

Tsotsos, J.K. (1990). Analyzing Vision at the Complexity Level. *Behavioral and Brain Sciences 13*, 423 - 445.

Walker, S. Stafford, P., & Davis, G. (2008). Ultra-rapid categorization requires visual attention: Scenes with multiple foreground objects. *Journal of Vision, 4*, 1-12.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception & Performance, 10*, 601-621.