Some Notes on the Control of Attention, its Modeling and Anticipatory Cognitive Computing

Sven Bertel

SFB/TR 8 Spatial Cognition, Universität Bremen, Germany
bertel@informatik.uni-bremen.de

Abstract

This paper is aimed at furthering discussions about the properties which computational cognitive models of attentional control should have if they are meant to be applied in interactive human-computer systems, namely for predicting human attention shifts across control levels. The paper discusses a number of issues of attentional control and implications that follow for computational modeling. It concludes by proposing the term *anticipatory cognitive computing* for those computational approaches that incorporate cognitive processing mechanisms found to exist in humans and that employ these mechanisms to generate hypothesis about imminent human cognitive and external actions.

Why a Modeling of Attention is Relevant for Good Human-Computer Interaction

In reasoning scenarios in which a human and a computational system jointly work on solving a common task, the system should not just respond to the human's actions but rather try to *anticipate* these. The corresponding aim is to achieve better joint performance of the duo (i.e., leading to higher quality of solutions, fewer errors, faster solution detection, and less frustration on the human part). This does not, but can, imply that the human's part in solving the task should be made a particularly easy one. It does imply, however, that the human's part should be made such as to best suit the working of his cognitive systems, that is, his reasoning capabilities, methods and styles.

Standard approaches to interface design often aim at creating interaction adequacy: the interfaces are optimized for interacting (i.e. for communicating) with a computational system. Striving for interaction adequacy is sufficient for many tasks. It is not where these tasks become cognitively demanding for the human and where the relationship between him and the computational system need to be more one of partners who collaboratively seek for a solution than that of a human operating a complex piece of machinery. Here, a *deeper* modeling is necessary, namely one that also aims at cognitive adequacy in reasoning (cf. the term *cognitive control* by Hollnagel & Caccibue, 1999).

To this end, Bertel (2005) argues for including predictive cognitive processing models within the computational reasoning system as to allow for it to predict specific cognitive parameters of its human counterpart (such as memory loads and spans, or foci of attention), problem parameters (e.g. regarding computational and representational complexity), and mental problem solving properties (such as preferences in mental model construction). These predictions should then be employed to modulate the actions of the computational system (i.e., to better adjust them to the human's cognitive processing capabilities, styles, and needs).

The set of applications for which it is sensible to establish such close human-computer collaboration in reasoning includes among others various tasks in spatial configuration and layout, many of which are in assistive or tutoring settings. Application domains that will likely increasingly require this sort of cognitive adequacy in partnership comprise architectural and product design, land-use planning, and many more (e.g. Bertel et al., in press). All of these domains have in common that the problems to be solved are at the same time complex and cognitively demanding and that a human reasoner can thus be expected to typically welcome some problem-specific computer-based assistance (e.g. Meniru et al., 2003, for computational needs in building design). Yet, with all, there exist reasons why the domain and the tasks cannot easily be completely described and why finding a formal specification is often just simply impossible. For example, with design problems, aspects that cannot be fully specified often relate to esthetics, to implicit knowledge of the designer, to questions of style or personal preferences. Consequently, one cannot address such problems by computational methods alone. In order to achieve satisfactory results, human reasoning needs to be an integral part of the overall problem solving process.

The Role of Attention and Control of Attention

Capturing and modeling the current attentional focus of the human reasoner is essential for the type of human-

---

computer collaboration intended: Given limited cognitive resources in memory and attention it is important to assess to which spatial and/or conceptual parts of the problem the human reasoner attends at a given moment and to which he does not. Corresponding data can then be fed into computational cognitive models and predictions across a number of control levels can be generated. These levels potentially range from oculomotor patterns (e.g. predicting next fixation targets in eye movements) to locations of foci of visual and non-visual attention, to aspects of the content and structure of constructed mental models, to cognitive style (both general and problem-specific), to mental and external (e.g. diagrammatic) problem solving strategies, and finally to next steps that are likely to occur within the problem solving process.

The corresponding spatial locations of attentional foci across different abstraction levels and corresponding control processes for these foci on the individual levels can serve as a cross-level binding mechanism that may eventually help gain new insights into and a better understanding of human (spatial) problem solving. The claim is that a better understanding will lead more accurate predictions and to better human-computer interaction.

Control of Attention

Following are a selection of fundamental issues on attention and the control of attention. The set is not meant to be exclusive, but should rather reflect those topics that seem to matter most for a computational modeling approach to attentional control which is targeted specifically at applications in interactive (reasoning) systems.

Two Concurrent Systems at Work

Neuroimaging and behavioral studies provide evidence for two kinds of attentional cognitive systems: One is essentially goal-directed (i.e. works top-down) and is involved in planning and executing spatial shifts of attention; the other is driven by perceptual influx (i.e. works bottom-up) (cf. Corbetta & Shulman, 2002). Both systems can be partially separated on a functional level and their implementations in the brain seem to be segregated for the most part (e.g. Corbetta et al., 2002).

In an ordinarily sighted human, the two systems are concurrently active and influence one another, for instance during visual search. The top-down system can effectively modulate the response of the bottom-up system for a sought object's location or visual features; this has been shown, for example, after spatial cueing (Posner, 1980) or after cueing for colors (Folk et al., 1992). The bottom-up system can in turn direct attention irrespective of the top-down system, especially after the occurrence of stimuli that are sudden or salient (e.g. Remington et al., 1992).

A function-oriented model of how attention is controlled – for instance, during spatial problem solving – thus cannot be monolithic but has to be composed at least of two competing subsystems for endogenous and for stimulus-induced activation.

Overlap of Attention with Working Memory

One defining characteristics of working memory is the ability to keep knowledge fragments activated in memory for a brief time period and to manipulate them also when no direct external stimulus is present. On an abstract, operational level, such abilities seem to require that one can attend to individual knowledge fragments and that the attentional focus can be shifted among fragments.

In fact, there is empirical support for a functional overlap in the mechanisms of selective spatial attention and those of spatial working memory, at least for rehearsal tasks (Awh & Jonides, 2001): Spatial rehearsal (that is, the maintenance of a spatial location in working memory) relies on spatial attention being directed to that location. Correspondingly, spatial memory performance declines when attention is directed away from the location or when shifts of attention are interrupted. This situation is somewhat analogous to one found with eye movements under imagery and retrieval conditions. The suppression of eye movement patterns associated with the memory of a particular spatial scene can lead to impaired long-term memory retrieval (Laeng & Teodorescu, 2002) and it seems rather likely that this would also be the case for non-spatial content associated to a spatial pattern (cf. the cognitive-perceptual binding as observed by Richardson & Spivey for retrieval of nonspatial information, 2000; or, more generally, Hommel, 2002, for evidence of integrated mental representations of spatial/nonspatial object information). Such connection hints at (a) a link between attentional control and control of visual focus and (b) at a potentially significant role of the original attentional foci to memory encoding, retrieval and/or manipulation in working memory (see also discussion below).

The notion of a functional overlap between spatial attention and working memory systems is further complemented by results from brain imaging studies indicating that the same cortical regions are recruited for attention shifting and spatial rehearsal (Corbetta et al., 2002).

The functional implications for a modeling of spatial attention are that attention cannot be addressed without a clear perspective on (spatial) working memory. As advocated by the predominant working memory models, attention systems need to be integrated with the individual working memory components (either in an extensively distributed manner as e.g. proposed by Cowan, 1999, or markedly more centralized as in Baddeley's model, e.g. 2002). However, it is likely that the spatial structure present during knowledge encoding has a much stronger influence on the encoding process, on the structure of mental representations formed, and on the working of retrieval processes than currently accounted for by these models. It seems plausible to assume that attentional patterns enacted during knowledge encoding can have a distinct influence on the control of attention during later knowledge retrieval and manipulation in working memory.

Spatio-Analogical Properties in Planning Shifts

The shifting of attention from one location to another seems to constitute a composed process. While the classical view advocates three distinct phases in attention shifting (namely, disengaging attention from a current attentional focus, moving attention to a target location or object, and engagement of the target; Posner et al., 1984), more current work suggests the existence of dissociable planning and execution stages which are at least partly serial (Hazlett & Woldorff, 2004). The latter scheme provides an interesting perspective on mental knowledge organization and on the structure of attentional control processes as only the planning stage has been found to depend on the spatial distance of shifting (i.e., the farther the attentional focus will be moved, the longer the planning phase).

It seems in line with general considerations on cognitive economy to assume that mental knowledge representations with spatio-analogical properties underlie the planning stage (i.e. a mental representation that bears a certain structural resemblance to physical space). In its analogicity, the planning of an attentional shift compares to other spatio-analogical mental processes that have been described, such as concerning the mental rotation of structured, three-dimensional objects (Shepard & Metzler, 1971). As a main implication for the modeling of attentional control, models of the planning of attention shifts across space have to take spatial properties of the shift (e.g. the distance between original and target locations) into account.

Attentional Control and Eye Movements

Shifts of attentional focus have frequently been attached to ‘a moving of the mind’s eye’. While such statement is clearly an exaggeration it may in fact not be very far from the truth: There exists broad psychological and functional anatomical evidence that link attentional processes to processes in eye movement control. Consequently, it has been argued that processes of attention and eye movements should be regarded as tightly connected and interdependent and that attentional shifts may even be fundamentally oculomotor in nature (Corbetta, 1998; Shepherd et al., 1986). This assumption is further corroborated by results from neural studies (e.g. Moore & Fallah, 2001).

Under normal conditions, attention and eye movements are synchronized and attentional and visual foci coincide on a common visual target (*overt attention*). During a fixation, however, the attentional focus can be dissociated from the eye coordinates and during an eye movement it can be moved even to an opposite direction (Posner, 1980; *covert attention*). Yet, there is also evidence that such dissociation is not complete as, for example, stimuli presented at fixation points show some spatial cueing effects regardless of whether the attentional focus coincides with a fixation point (cf. Shepherd et al., 1986). By recording and analyzing a person’s eye movements one can consequently get a robust estimate of the current focus of attention. As a corollary, models that predict or explain eye fixations cannot be segregated from models that predict or explain shifts of attention.

It has been argued above that attentional patterns may form part of the memory of a spatial scene or configuration. This is also true for eye movements (i.e. with regard to the sequences of fixations; e.g. Noton & Stark, 1971). The original fixation sequences can not only influence eye movements during later scene inspections or during memory access; their suppression can also lead to impaired retrieval suggesting that eye movements play a functional role for mental (spatial) reasoning (Laeng & Teodorescu, 2002). The function of the original oculomotor patterns is such that they have been even found to be reproduced under imagery conditions (Brandt & Stark, 1997), hinting at a fundamental organizational principle. This impression is further substantiated by the finding that – after first listening to verbal descriptions of a story – subjects’ eye movements during later recall reflect much of the spatial relations contained in the story (Johansson et al., 2005). Possible interpretations would be that subjects were simply recalling a mental scene they had constructed while listening to the story and that (during later recall) they either reenacted the original eye movements related to the construction process or mentally inspected the scene during which mental attentional shifts also induced eye movements. One should keep in mind, however, that internal shifts of attention do not necessarily entail eye movements, particularly when the object observed (or, likely, also imagined) is small and lies entirely within the current visual focus (or within the focus of attention under imagery) (e.g. Noton & Stark, 1971).

Last, eye movements have also been found to reflect cognitive processes involved in diagrammatic problem solving, hinting at a relation to higher-level cognitive control principles (i.e., to issues of control of focus during problem solving: *Which parts of the problem are attended to at which moment? Which are the strategies employed? Which solution models are being mentally constructed, and which ones aren’t?* etc.; Bertel, 2006; cf. also Knoblich et al., 2001). In particular with diagrammatic problems, sequences of visual attention shifts (i.e. partial *scanpaths*) often carry characteristic meaning as their mental representation can be bound to that of specific substructures of the problem. Conversely, the direction of visual attention to specific parts of a diagram by means of visual cues can lead to changes in attentional shifts and/or eye movements, possibly to changes in problem solving foci, and also to improved performance (Grant & Spivey, 2003). At least for problems in spatial or diagrammatic reasoning it seems feasible to use eye movements as an input vector for a dynamic modeling of attentional shifts and to gradually abstract from there on to a dynamic model of focus in problem solving.

Models of Attention and Control of Attention

The basic questions are these: How should attentional control be conceptualized and computationally modeled on and across cognitive processing levels? And: how do we get data on current attentional statuses? Following are a non-exhaustive collection of issues and approaches related to these two questions.

Levels of Modeling

It is obviously an important issue, which levels across the perceptual-cognitive or low-level-high-level spectra (in this case, between the oculomotor processing and the problem solving strategies) one should address in modeling. It directly relates to the type of interplay between the bottom-up and top-down attention control systems to be included. Approaches that emphasize bottom-up processing and perceptual saliency (e.g. Itti & Koch, 2001) are well suited for explaining why attention gets grabbed by a certain stimulus, sometimes regardless of input from higher levels (as in the case of a fire alarm going off). Hybrid approaches (e.g. Schill et al., 2001) can explain why saccades occur given both a stimulus and prior knowledge or expectations of a visual scene. More laterally organized approaches try to primarily relate attentional focus and control principles across representational systems on same or neighboring levels of abstraction rather than across these (for instance, regarding basic spatial organization principles and processing mechanisms in diagrams, mental images and mental models; Engel et al., 2005). Other, more application-oriented approaches target cross-modal aspects and implications that can be drawn for the design of user interfaces from behavioral research (e.g. Stanney et al., 2004).

The field thus is a vast one, and to incorporate in a model all aspects of control across abstraction levels that likely in some way or the other influence the mental attention shifting mechanisms does not fall short of creating a truly unified theory of cognition. To be practicable, any comprehensive approach must therefore likely abstract from many issues and concentrate on a few, and it naturally depends on the modeling goals that one pursues:

- *Start small and problem-specific* can be a good guideline for control of attention models that will eventually be employed in specific application settings. With respect to the target domain of collaborative human-computer spatial reasoning, it seems that robust models of bottom-up, perceptual processes may be sufficient for many contexts. When the goal is to distinguish between high-level problem-solving attributes that a human reasoner may employ rather than to give an accurate predictive account of the next saccadic eye movement, the focus of the model should be on specific (i.e. distinctive) qualitative and quantitative aspect of lower level processes and feed the gained data into more abstract levels. For example, in the diagrammatic problem solving paradigm sketched in Bertel (2006), human preferences for different solution models were distinguished by analyzing eye movement patterns over different temporal intervals and spatial regions. While the individual fixation may not be as important for such contexts fixation patterns clearly are.
- *Be modular and cross-modal* is likely a motto of choice for models that are primarily aimed at understanding human cognition. Models of the first type may be excellent engineering solutions to human-computer collaborative reasoning for a certain domain; the downside is that they probably do not scale up well to larger or different domains. Where then lie the alternatives? Clearly, they are neither in an exhaustive modeling of all levels and cross-level interactions relevant to attentional control, nor in constructing all models of higher level processes exclusively bottom-up from processes on perceptual levels. Both approaches would be too broad and unfocused, if for different reasons. Instead, it seems more promising to construct a variety of domain-specific models which are linked up by way of common underlying representational and processing principles (i.e. in the sprit of ideas on ‘distributed, yet structurally similar processes’ by Cowan, 1999, Nobre et al., 2004, or Engel et al., 2005). In such a paradigm, larger models get constructed by linking up collections of smaller ones, establishing interconnections between these, and adapting or reconstructing smaller models when they cannot be integrated. While this permits a somewhat organic growth of the overall models and does allow for boot-strapping techniques to be employed, the results will likely be somewhat unstructured or even chaotic from a system engineering point of view.

General Modeling and Individual Differences

Regarding issues of generality versus inter-individual differences in modeling it seems practicable to assume a general cognitive architecture of attentional processing for all human reasoners and to adjust this architecture to individual requirements. For example, one can adjust a general model based on influx of data on attention and the richness of the data available (data-dependency), based on detected personal cognitive styles (e.g. on the visualizer/verbalizer scale; Kozhevnikov et al., 2002), on personal mental capacities, developed strategies, performance level, etc. (cognizer-dependency), based on the problem domain and task (task-dependency), or based on other contexts.

How Can We Get Data?

Getting live data on current foci of visual attention seems to be crucial for running computational models of attentional control in real-world application contexts (see the discussion on eye tracking and attention, above). It therefore is a natural choice, at least for spatial (e.g. diagrammatic) reasoning contexts, to employ eye tracking techniques and to propagate the gathered data vertically or horizontally through to other abstraction levels of attention models. Preferably, such data can be complemented by and integrated with data from other modalities, for example originating from drawing actions, speech, or gestures produced at the same time.

Anticipatory Cognitive Computing

Anticipatory computing addresses the computational relations that exist between actions on the one hand and expectations, predictions, and forecasts, on the other (cf. Nadin, 2000). *Cognitive computing* denotes artificial problem solving techniques that at least in part mimic techniques employed by natural cognitive systems (e.g. by humans; cf. Erickson, 1993). Based on the discussions in this paper and the general approach that emerges from these in human-computer collaboration and computational cognitive modeling, I propose to use the term *anticipatory cognitive computing (ACC)*. It employs problem solving techniques, structural principles and cognitive processing mechanisms derived from studying human cognitive processing, it creates computational cognitive models based on these techniques, principles and mechanisms, and it then applies the resulting models to generate cognitively-founded anticipatory behavior (i.e. in term of anticipation of imminent human cognitive processing and actions) in an artificial cognitive system.

With respect to modeling control of attention, an ACC system should include actions generated from or modulated by predictions on a human's focus of attention, and visual or other-modal focus. It should be able to integrate a live influx of behavioral data on attention recorded from a human reasoner. Depending on the purpose and modeling paradigm or approach chosen, it will also likely contain a collection of inner states that model the production of overt and covert attentional shifts and the significance of these changes to solving the task at hand.

References

Awh, E. and Jonides, J. 2001. Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, 5(3), 119–126.

Baddeley, A. 2002. Fractionating the central executive. In: T. D. Stuss, and T. R Knight (Eds), *Principles of Frontal Lobe Function* (pp. 246–260). New York: Oxford University Press.

Bertel, S. 2005. Show me how you act on a diagram and I'll tell you what you think. In *Reasoning with Mental and External Diagrams: Computational Modeling and Spatial Assistance. Papers from the 2005 AAAI Spring Symposium (SS-05-06).* AAAI; Menlo Park, CA.

Bertel, S. 2006. Visual focus in computer-assisted diagrammatic reasoning. In D. Barker-Plummer, R. Cox, and N. Swoboda (Eds), *Diagrammatic Representation and Inference, Proc. 4th International Conference Diagrams 2006* (pp. 241–244). LNAI, Berlin: Springer.

Bertel, S., Vrachliotis, G., and Freksa, C. In press. Aspect-oriented building design: Towards computer-aided approaches to solving spatial constraints in architecture. In G. Allen (Ed.), *Applied Spatial Cognition: From Research to Cognitive Technology*, Mahwah, NJ: Lawrence Erlbaum Associates.

Brandt, A. S. and Stark, W. L. 1997. Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience*, 9, 27–38.

Corbetta, M. 1998. Frontoparietal cortical networks for directing attention and the eye to visual locations: identical, independent, or overlapping neural systems? *Proc. Natl. Acad. Sci. U.S.A.*, 95, 831–838.

Corbetta M., Kincade J. M., and Shulman G. L. 2002. Neural systems for visual orienting and their relationships to spatial working memory. *Journal of Cognitive Neuroscience*, 14, 508–23.

Corbetta, M. and Shulman, G. L. 2002. Control of goal-driven and stimulus driven attention in the brain. *Nature Reviews Neuroscience* 3, 201–15.

Cowan, N. 1999. An embedded-process model of working memory. In: P. Shah, and A. Miyake (Eds), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 62–101). New York: Cambridge University Press.

Engel, D., Bertel, S., and Barkowsky, T. 2005. Spatial Principles in Control of Focus in Reasoning with Mental Representations, Images, and Diagrams. C. Freksa et al. (Eds.), *Spatial Cognition IV.* Springer, Berlin.

Erickson, J. 1993. What is cognitive computing? *Dr. Dobb's Journal*, 1993.

Folk, C. L., Remington, R. W., and Johnston, J. C. 1992. Involuntary covert orienting is contingent on attentional control setting. *Journal of Experimental Psychology: Human Perception and Performance,* 18(4)*,*1030–1044.

Grant, R. E. and Spivey, J. M. 2003. Eye movements and problem solving: guiding attention guides thought. *Psychological Science*, 14(5), 462–466.

Hazlett, J. C. and Woldorff, G. M. 2004. Mechanisms of moving the mind's eye: Planning and execution of spatial shifts of attention. *Journal of Cognitive Neuroscience*, 16(5), 742–750.

Hollnagel, E. and Cacciabue, P. C. 1999. Cognition, Technology & Work: An introduction. *Cognition, Technology & Work*, 1, 1–6.

Hommel, B. 2002. Responding to object files: Automatic integration of spatial information revealed by stimulus–response compatibility effects. *Quarterly Journal of Experimental Psychology*, 55A(2), 567–580.

Itti, L. and Koch, C. 2001. Computational modelling of visual attention. *Nat. Rev. Neurosci.*, 2, 194–203.

Johansson, R., Holsanova, J., and Holmqvist, K. 2005. What Do Eye Movements Reveal About Mental Imagery? Evidence From Visual And Verbal Elicitations. In Bara, B. G., Barsalou, L., and Bucciarelli, M. (Eds.), *Proc. 27th Annual Conference of the Cognitive Science Society* (pp. 1054–1059). Mahwah, NJ: Erlbaum.

Knoblich, G., Ohlsson, S., and Raney, E. G. 2001. An eye movement study of insight problem solving. *Memory & Cognition*, 29(7), 1000–1009.

Kozhevnikov, M., Hegarty, M., and Mayer, R. E. 2002. Revisiting the visualizer-verbalizer dimension: Evidence for two types of visualizers. *Cognition & Instruction*, *20*, 47–78.

Laeng, B. and Teodorescu, D. 2002. Eye scanpaths during visual imagery reenact those of perception of the same visual scene. *Cognitive Science*, 26, 207–231.

Meniru, K., Rivard, H., and Bédard, C. 2003. Specifications for computer-aided conceptual building design. *Design Studies*, 24, 51–71.

Moore, T. and Fallah, M. 2001. Control of eye movements and spatial attention. *Proc. Nat. Acad. Sci. USA*, 98, 1273–1276.

Nadin, M. 2000. Anticipatory Computing. *Ubiquity - The ACM IT Magazine and Forum, Views*, 1(40).

Nobre, A. C., Coull, J. T., Maquet, P., Frith, C. D., Vandenberghe, R., and Mesulam, M.M. 2004. Orienting attention to locations in perceptual versus mental representations. *Journal of Cognitive Neuroscience*, 16(3), 363–373.

Noton, D. and Stark, L. 1971. Eye movements and visual perception, *Scientific American*, 224, 34–43.

Posner, M. I. 1980. Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3–25.

Posner, M. I., Walker, J. A., Friedrich, F. J., and Rafal, R. D. 1984. Effects of parietal injury on covert orienting of attention. *Journal of Neuroscience*, 4(7), 1863–1874.

Remington, R. W., Johnston, J. C., and Yantis, S. 1992. Involuntary attentional capture by abrupt onsets. *Perception and Psychophysics*, 51(3), 279–290.

Richardson, C. D. and Spivey, J. M. 2000. Representation, space and Hollywood Squares: looking at things that aren't there anymore. *Cognition*, 76, 269–295.

Schill, K., Umkehrer, E., Beinlich, S., Krieger, G., and Zetzsche, C. 2001**.** Scene analysis with saccadic eye movements: top-down and bottom-up modelling, *J. of Electronic Imaging*, 10(1), 152–160.

Shepard, N. R. and Metzler, J. 1971. Mental rotation of three-dimensional objects. *Science*, 171, 701–703.

Shepherd, M., Findlay, J. M., and Hockey, R. J. 1986. The relationship between eye movements and spatial attention. *Quarterly J. of Experimental Psychology,* 38A, 475–491.

Stanney, K., Samman, S., Reeves, L., Hale, K. S., Buff, W., Bowers, C., Goldiez, B., Nicholson, D., and Lackey, S. 2004. A paradigm shift in interactive computing. *International Journal of Human-Computer Interaction*, 17(2), 229–257.