Gem Stapleton John Howse John Lee (Eds.)

Diagrammatic Representation and Inference

5th International Conference, Diagrams 2008
Herrsching, Germany, September 19-21, 2008
Proceedings

With 13 Color Figures

Springer

Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada
Jörg Siekmann, University of Saarland, Saarbrücken, Germany
Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

Volume Editors

Gem Stapleton
Computing, Mathematical and Information Sciences
University of Brighton
Brighton, UK
E-mail: G.E.Stapleton@brighton.ac.uk

John Howse
Computing, Mathematical and Information Sciences
University of Brighton
Brighton, UK
E-mail: John.Howse@brighton.ac.uk

John Lee
Human Communication Research Centre
University of Edinburgh
Informatics Forum
Edinburgh, Scotland, UK
E-mail: J.Lee@ed.ac.uk

Library of Congress Control Number: 2008934909

CR Subject Classification (1998): I.2, D.1.7, G.2, H.5, J.4, J.5

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN	0302-9743
ISBN-10	3-540-87729-0 Springer Berlin Heidelberg New York
ISBN-13	978-3-540-87729-5 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12525874 06/3180 5 4 3 2 1 0

Multimodal Comprehension of Graphics with Textual Annotations: The Role of Graphical Means Relating Annotations and Graph Lines

Cengiz Acarturk[1,*], Christopher Habel[1], and Kursat Cagiltay[2]

[1] University of Hamburg, Computer Science, Vogt-Koelln-Str. 30,
22527 Hamburg, Germany
{acarturk,habel}@informatik.uni-hamburg.de
[2] Middle East Technical University, Computer Education and Instructional Technology
06531 Ankara, Turkey
kursat@metu.edu.tr

Abstract. Graphs are often accompanied by text, i.e. linguistically coded information, augmenting the information presented diagrammatically. Thus, graph comprehension by humans often constitutes comprehension and integration of information provided by different representational modalities, namely graphical elements and verbal constituents. In this study we focus on textual annotations to line graphs providing information about events, processes and their temporal properties as well as temporal relations about the events and processes in question. We present results of an experimental investigation on parameters which influence subject's interpretations concerning the temporal properties of the annotated events and on eye movement behavior. In particular, we discuss the role of graph shape and the role of graphical means for relating textual annotations and determined parts of the graph line.

Keywords: text-graphics comprehension; annotations; line graphs; temporal relations.

1 Introduction

Multimodal documents combining text and pictorial representation such as newspaper articles, educational material and scientific papers are wide-spread in print media as well as in electronic media.[1] Comprehension of multimodal documents is based on almost automatically performed cognitive processes underlying the integration of information provided by the different modalities. Researchers from different disciplines investigated multimodal documents of different types in different domains, [2], [3], [4] among many others. Nevertheless, the research on cognitive mechanisms underlying multimodal integration is currently in a premature state due to abundant

* The research reported in this paper has been partially supported by DFG (German Science Foundation) in ITRG 1247 'Cross-modal Interaction in Natural and Artificial Cognitive Systems' (CINACS). We thank three anonymous reviewers for their helpful comments.

[1] In this paper, we use the term 'modality' as shorthand for 'representational modality' [1].

G. Stapleton, J. Howse, and J. Lee (Eds.): Diagrams 2008, LNAI 5223, pp. 335–343, 2008.

possible variations of the external representations and the modes of communication (e.g., speech vs. written text). In the present paper we focus on a specific type of pictorial representations, namely on *diagrams*—in particular *line graphs*—and on comprehension of diagrams in the context of text. The research on *cognitive* (rather than *perceptual*) processes in graph comprehension is scarce [5] in the last decades with some exceptions [6], [7], [8]. Furthermore, multimodal comprehension of graph-text elements has seldom been in the main focus of the research so far.

In contrast to *pictures* (or *images*) diagrams possess internal syntactic structures in the sense of representational formats [9]. Thus the syntactic analysis of a graph can be exploited by succeeding processes of semantic and pragmatic analyses in graph comprehension [7], [10], [11]. From a linguistic point of view, the process of *referring*, which is constituted by a *referential expression* that refers to an *entity* of the domain of discourse, is the core of comprehension. Based on this, *co-reference*, the backbone of text coherence has to be established by speaker and hearer employing internal—conceptual—representations, which mediate between language and the domain of discourse. In processing text-diagram documents, in which both modalities have to contribute to a common conceptual representation, additional types of reference and co-reference relations have to be distinguished. Foremost, there exist corresponding referential relations (reference links) between graphical entities and entities in the domain of discourse. Furthermore, there exist referential links between linguistic and graphical entities.[2] Beyond this—traditionally discussed—type of text-diagram multimodality there exists a second kind, the diagram-internal multimodality of graphical representations and diagram-internal text, e.g. labeling of axes, annotation to graph lines etc. Figure 1 shows that both types of multimodality can be involved in multimodal comprehension. Therefore it is important to investigate the role of annotating textual elements in the graph region of a text-diagram document.

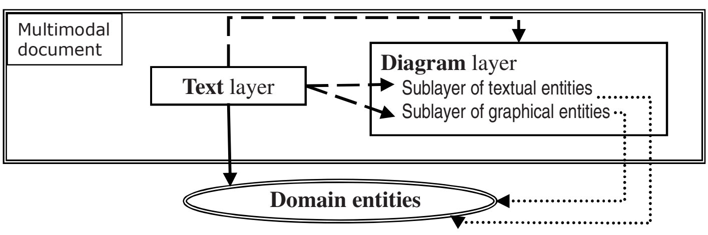

Fig. 1. Reference links between layers and sub-layers of text-diagram documents

The graph region includes the graph (proper) and graph-related text information (graph title, annotations etc.), and usually is separated by a frame from the rest of the document (e.g. paragraphs). We propose that annotated textual elements serve the purpose of bridging the two representational modalities in graph-text documents [13].

[2] In [12] we discussed the concept of *text-graphic coherence* based on an analysis of different types of referential links. Furthermore, we described how the interaction between information graphics and language is mediated by common conceptual representations.

If annotations are available contiguous to the graph, separately constructed representations of the text and the graph can be connected via these constituents. If not, integration of modalities is achieved with further cognitive effort of encoding spatially represented information on the graph and constructing co-reference relations between the paragraphs and the graph. Our purpose in the present study is to investigate the diagram-internal multimodality of graphical elements and diagram-internal text, i.e. textual annotations. We exemplify this topic focusing on line graphs concerning sequences of events/processes and describing temporal properties of these events and processes (see Figure 2). For annotated line graphs, the content provided distributively by graphical elements and verbal annotations has to be integrated via co-reference relations to reach coherent conceptual representations. The construction of co-reference links can be induced by spatial contiguity between annotations and the graph line or prominent parts of the graph line (e.g. "Famine 1930" or "1948 lake freeze-over") or by explicit 'pointers', called 'annotation icons', (e.g. "Fire 1936")[3].

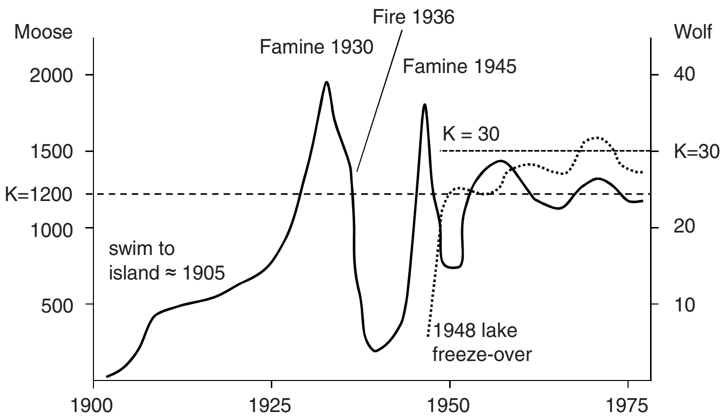

Fig. 2. A sample annotated line graph: Isles Royale Moose/Wolf Progression[4]

Accordingly, the following research questions were central to the present study: How do graph format (i.e. straight- vs. curved-shape) and annotation icon affect interpretations concerning the temporal properties of the annotated event? How do the salient points change under the presence of annotation text and annotation icon?

Methodologically, compared to research on eye movement control in reading, there are few studies investigating eye movement characteristics in multimodal documents

[3] In this study, we use the term 'annotation' to mean the combination of textual elements (e.g. phrases) and a connecting icon, for example a vertical line which on the one end points the textual element, on the other end points a specific location on the graph. The term 'annotation icon' is used to mean this vertical line. The term 'annotation text' is used to mean the verbal constituents of the annotation.

[4] Figure 2 is based on a line graph used in N.C. Heywood's course material in Biogeography (http://www.uwsp.edu/geo/faculty/heywood/Geog358/Population/Populate2.htm). Fig. 2 goes back to Harris, A. and Tuttle, E. Geology of National Parks. ISBN 0-8403-2810-9.

[14]. The underlying assumptions in these studies, as well as in the present study, are based on the *eye-mind hypothesis* [15]; see [16] for objections to the eye-mind hypothesis in its proposed form.

2 The Experiment

Subjects. A total of 36 subjects (mean age 22.8, SD = 2.49) were paid to participate in the experiment. The subjects were either undergraduate or graduate students in different general academic areas at the Middle East Technical University. All subjects were native speakers of Turkish, which was the language of the experiment.

Materials and Design. Simple line graphs that present the change of a domain value in time were prepared by the experimenter. The graphs included an annotation that includes information about an event's occurrence (see Figure 3).

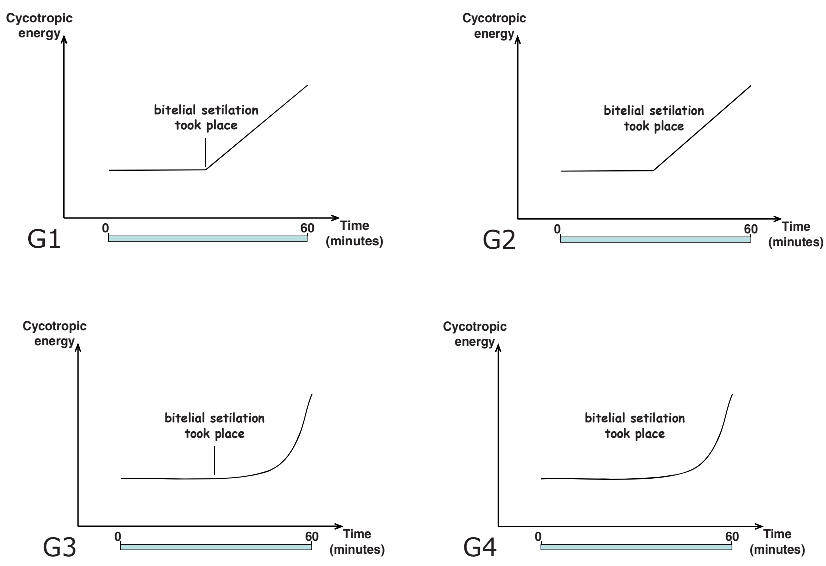

Fig. 3. The four graphs show sample material for the four experimental conditions G1 to G4 (translated to English by the first author; fonts in the figure were changed for better visibility)

The design was 2x2, with two independent parameters (two within-subjects parameters), and one dependent parameter. The first independent parameter was the *graph format*. The graph was either a straight-shaped line graph (G1 and G2) or curve-shaped line graph (G3 and G4). The second independent parameter was the *annotation icon*. The graph either included an annotation icon (G1 and G3) or did not include an annotation icon (G2 and G4). As a result, each subject was presented four experimental conditions. The dependent parameter was subjects' reports of interpretations concerning the *duration* of the annotated event, explained below. The order of presentation of the conditions was randomized. In addition to the four experimental conditions, two graphs with straight- and curve-shaped lines without annotations were presented to the subjects without a specific given task (C1 and C2). They were presented for the purpose of comparison with the experimental conditions (G1 to G4). The domain was not evaluated as an independent parameter in this study. The domain value labels, as well as the annotated events were prepared for four different fictional domains. Subjects were informed that the graphs were excerpted from lecture notes in medicine, which was an unfamiliar domain for the participants of the study. Eye-tracking data were recorded by a 50 Hz. Tobii 1750 EyeTracker.

Procedure. The subjects attended the experiment in single sessions. After the preliminary information about statistical information graphics with samples, a practice session was presented to explain the task and the use of the response scale. In the experiment session, in each screen, subjects investigated the graph and reported their interpretation by clicking on the response scale, which was given below the x-axis in the same screen (see Figure 3). The response scale was a colored horizontal bar extending from the beginning to the end of the horizontal time axis of the graph. If the subject interpreted the event as point-like, i.e. the event occurred in a specific point in time then he/she clicked the time when the event happened, on the response scale with the mouse (i.e. point interpretation). If the subject interpreted the event as durative, i.e. the event occurred in a time interval rather than a specific point in time then he/she clicked the time when the event started and the time when the event ended, on the response scale (i.e. interval interpretation). The experiment was self-paced, and took a total of approximately 10 minutes to complete.

3 Results

The distribution of the number of subjects who made the *point* and *interval* interpretations for the temporal properties of the annotated event showed that the main determinant was the *graph format*. The presence or absence of the *annotation icon* had a marginal effect. A Cochran test was conducted to evaluate differences between related proportions. The test was significant, χ^2 (3, N = 36) = 20.47, $p < .01$, Kendall coefficient of concordance was .19. Follow-up pairwise comparisons were conducted using a McNemar's test. The results showed that the number of subjects who made point interpretation was significantly higher in conditions G1 and G2 than the number of subjects who made point interpretation in conditions G3 and G4. Correspondingly, the number of subjects who made interval interpretation (almost half of the subjects) was significantly higher in conditions G3 and G4 than the number of subjects who made interval interpretation in conditions G1 and G2.

Further analysis of the effect of the annotation icon as well as the graph format was investigated with the analysis of eye movement parameters, namely fixation count, gaze time and fixation duration. For the analysis, the region covered by the graph line was divided into 15 rectangle AOIs (Area of Interest), namely *AOI 1* to *AOI 15* (Figure 4). The same AOI template was used to evaluate all experimental conditions.

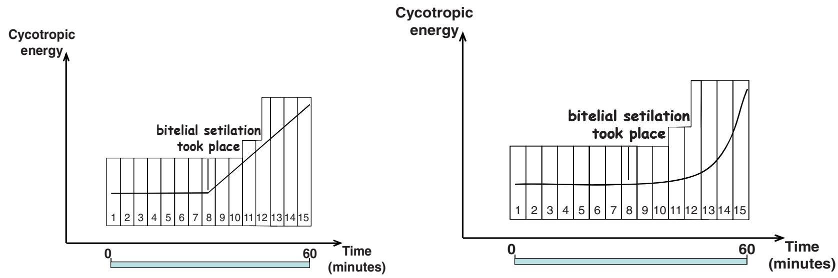

Fig. 4. The specified AOIs (Area of Interest) for the analysis of eye movement parameters

Mean fixation counts on the AOIs were calculated, and z-score normalization was used for the analysis. A within-subjects repeated measures ANOVA was conducted with the factors being the conditions and mean fixation counts on the fifteen AOIs. The results indicated a significant condition effect, Wilks's $\Lambda = .18$, $F(5, 31) = 28.95$, $p < .01$, a significant AOI effect, Wilks's $\Lambda = .04$, $F(14, 22) = 35.53$, $p < .01$, and a significant interaction between the conditions and the AOIs. The distribution of the mean fixation counts for C1 (the straight graph without annotations) and C2 (the curved graph without annotations), shown in the left part of Figure 5, reveals information about visually/informationally salient regions on the graph lines. In the C1 graph the salient region was AOI 8, whereas in the C2 it was AOI 12.

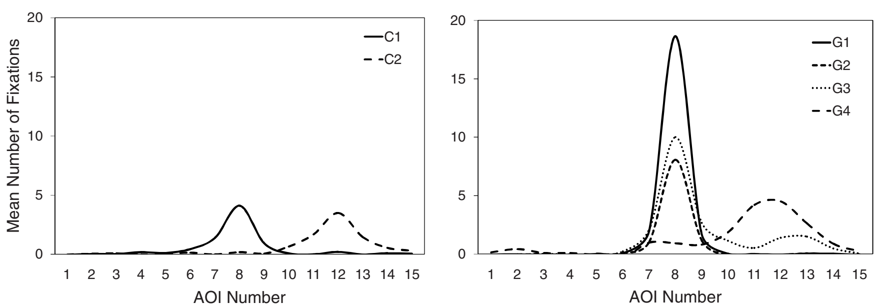

Fig. 5. The distribution of mean number of fixations (i.e. fixation counts) on the AOI 1 to AOI 15 in C1 and C2 conditions (on the left) and G1, G2, G3, and G4 conditions (on the right)

How does the distribution change in the presence of annotation text? This corresponds to the conditions G2 and G4, in other words the two annotated graphs with different graph format and without annotation icon. The results (the right part of Figure 5) revealed a similar distribution to the C1-C2 distribution. The salient region in the G2 condition was AOI 8, whereas the salient regions in the G4 condition were AOI 11 and AOI 12. Furthermore, the comparison of mean fixation counts for the C1-C2 conditions and the G2-G4 conditions shows that in G2 and G4, the addition of the annotation text resulted in an overall increase in the number of eye fixations on the AOI 8, which was the region below the annotation text.

How does the distribution change in the presence of an annotation icon, as well as the annotation text? This corresponds to the conditions G1 and G3, in other words the two annotated graphs with different graph format and with the annotation icon. The results showed that the distribution of mean fixation counts in the G1 and G3 conditions was different than the distributions in the previous conditions. The comparison of mean fixation count distributions for the conditions G1 and G3 shows that the salient region was AOI 8 for both conditions. In other words, especially for the G3 condition, the visually/informationally salient region was shifted from AOI 11 and AOI 12 (in the condition G4) to AOI 8 (in the condition G3) with the addition of the annotation icon.[5]

The analysis of gaze time values on the previously specified AOIs revealed similar distributions to the ones for mean fixation counts.

The results for mean fixation durations were calculated for the AOIs that had an average number of fixations of one or greater than one. Accordingly, mean fixation durations were calculated for AOI 7 to AOI 9 in G1 and G2 conditions and for AOI 7 to AOI 13 in G3 and G4. A univariate ANOVA test was conducted for the analysis of the differences between conditions G3 and G4. The results for the ANOVA indicated a significant condition effect, $F(1, 296) = 6.09$, $p < .05$ and a significant AOI effect, $F(6, 296) = 3.15$, $p < .01$. Important from the focus of this study is that mean fixations in the G4 condition were longer than the ones in the G3 condition.

4 Discussion

The subject's reports of interpretations concerning the temporal properties of the annotated events show that the graph format, rather than the presence or absence of the annotation icon, is the main determinant for the temporal properties of the annotated events. Most of the subjects reported that the annotated event took place at a specific point in time in the straight-graph conditions. On the other hand, almost half of the subjects reported that the annotated event took place in a time interval in the curved graph conditions. Nevertheless, the role of the annotation icon on subjects' reports is not significant between the conditions.

Further analysis of eye movement parameters reveals more detailed information for the differences between the conditions and the effect of the presence or absence of an annotation icon. The results show that on the one hand, the addition of the annotation texts to the graphs does not reveal major changes in the distribution of average number of fixations and gaze time values on the previously specified AOIs, compared to the distributions on the non-annotated graphs. This implies that the annotation text, without the annotation icon does not strongly affect the visually/informationally salient regions on the graphs. On the other hand, mean fixation counts and gaze time values increase with the addition of the annotation text. The increase in fixation counts and gaze time values on the graph line points to subjects' effort for the integration of information provided by the graphical elements and the annotation text.

[5] Whether this shift is stable, or is an artifact due to the very slight change at the beginning of the curves (*inflections)* used in this experiment—as suspected by two reviewers—will be investigated in a future study. *Inflections* have been found to be the second mostly used segmentation points after *negative minima* [17], [18]. We argue here, that annotation icons belong to those *top down factors* which interact with geometric factors in segmentation [18].

Furthermore, the addition of an annotation icon (together with the addition of the annotation text) results in major changes in the distribution of mean fixation counts and in the distribution of average gaze time values in the curved graphs. In other words, the presence of annotation icons shifts the visually/informationally salient points in the curved graphs. This is expected, since the addition of the new graphical element, namely the annotation icon, attracts subjects' attention to this region. In addition, average number of fixations and gaze time values further increase with the addition of the annotation icon. This increase may imply subjects' further effort to integrate the information provided by the annotation text and the annotation icon, as well as the annotation icon and the *relevant part of the graph*.

More important from the perspective of this study is that the subjects experienced difficulties in determining the *relevant part of the graph* in the absence of the annotation icon in the curved graphs. The results of the analysis for mean fixation durations support this idea showing that the absence of the annotation icon in the curved graphs results in longer fixations on the salient regions of the graph.

5 Conclusions and Future Work

Comprehension of multimodal documents includes the construction of co-reference relations between the graph (proper) and text elements (e.g. annotations) in the graph region, as well as the construction of co-reference relations between the graph region and the main text (e.g. paragraphs) of the document. From the perspective of automated generation of graph-text documents, investigation of multimodal integration at both levels is necessary for the design of easily comprehended documents and graphs by humans. Poorly designed diagrams may create misconceptions, deceive or confuse an issue, do not ease the comprehender's task and hinder comprehension and learning. Nevertheless, in the current state of the art, the design and use of annotations is based on the experience and practice of the designers of graphs and multimodal documents, rather than theory, guidelines or systematic empirical research. Furthermore, on the application side, recent data visualization components of popular statistical and mathematical software programs offer limited capacity for annotation design and generation for graphs. Preparing effective diagrams requires both practice and also evidences from empirical research studies.

In this study, we investigated how the graphical elements (i.e. the graph format, as well as the absence or presence of an annotation icon) affect subjects' interpretations concerning the temporal properties of events annotated by textual elements, and eye movement characteristics during multimodal graph comprehension. In the future, we will investigate the role of additional graphical means (e.g. arrows as well as lines, textbox etc.) for annotations by experiments with human subjects, as well as by corpus studies.

References

1. Bernsen, N.O.: Foundations of Multimodal Representations: A Taxonomy of Representational Modalities. Interacting with Computers 6, 347–371 (1994)
2. Hegarty, M.: The Mechanics of Comprehension and Comprehension of Mechanics. In: Rayner, K. (ed.) Eye Movements and Visual Cognition: Scene Perception and Reading, pp. 428–443. Springer, New York (1992)

3. Butcher, K.R.: Learning from Text with Diagrams: Promoting Mental Model Development and Inference Generation. Journal of Educational Psychology 98, 182–197 (2006)
4. Narayanan, N.H., Hegarty, M.: On Designing Comprehensible Interactive Hypermedia Manuals. International Journal of Human Computer Studies 48, 267–301 (1998)
5. Scaife, M., Rogers, Y.: External Cognition: How Do Graphical Representations Work? International Journal of Human Computer Studies 45, 185–213 (1996)
6. Peebles, D.J., Cheng, P.C.-H.: Extending Task Analytic Models of Graph-based Reasoning: A Cognitive Model of Problem Solving with Cartesian Graphs in ACT-R/PM. Cognitive Systems research 3, 77–86 (2002)
7. Pinker, S.: A Theory of Graph Comprehension. In: Freedle, R. (ed.) Artificial intelligence and the Future of Testing, pp. 73–126. Erlbaum, Hillsdale (1990)
8. Mautone, P.D., Mayer, R.E.: Cognitive Aids for Guiding Graph Comprehension. Journal of Educational Psychology 99, 640–652 (2007)
9. Kosslyn, S.M.: Image and Mind. Harvard University Press, Cambridge (1980)
10. Kosslyn, S.M.: Understanding Charts and Graphs. Applied Cognitive Psychology 3, 185–226 (1989)
11. Tversky, B.: Semantics, Syntax, and Pragmatics of Graphics. In: Holmqvist, K., Ericsson, Y. (eds.) Language and Visualisation, pp. 141–158. Lund University Press, Lund (2004)
12. Habel, C., Acarturk, C.: On Reciprocal Improvement in Multimodal Generation: Co-reference by Text and Information Graphics. In: van der Sluis, I., Theune, M., Reiter, E., Krahmer, E. (eds.) Workshop on Multimodal Output Generation (MOG 2007), Aberdeen, United Kingdom, pp. 69–80 (2007)
13. Acarturk, C., Habel, C., Cagiltay, K., Alacam, O.: Multimodal Comprehension of Language and Graphics: Graphs with and without Annotations (in press, accepted for publication in Journal of Eye Movement Research)
14. Underwood, G., Jebbett, L., Roberts, K.: Inspecting Pictures for Information to Verify a Sentence: Eye Movements in General Encoding and in Focused search. The Quarterly Journal of Experimental Psychology 57A, 165–182 (2004)
15. Just, M.A., Carpenter, P.A.: A Theory of Reading: From Eye Fixations to Comprehension. Psychological Review 87, 329–354 (1980)
16. Anderson, J.R., Bothell, D., Douglass, S.: Eye Movements Do Not Reflect Retrieval Processes: Limits of the Eye-mind Hypothesis. Psychological Science 15, 225–231 (2004)
17. Cohen, E., Singh, M.: Geometric Determinants of Shape Segmentation: Tests Using Segment Identification. Vision Research 47, 2825–2840 (2007)
18. De Winter, J., Wagemans, J.: Segmentation of Object Outlines into Parts: A Large-scale Integrative Study. Cognition 99, 275–325 (2006)