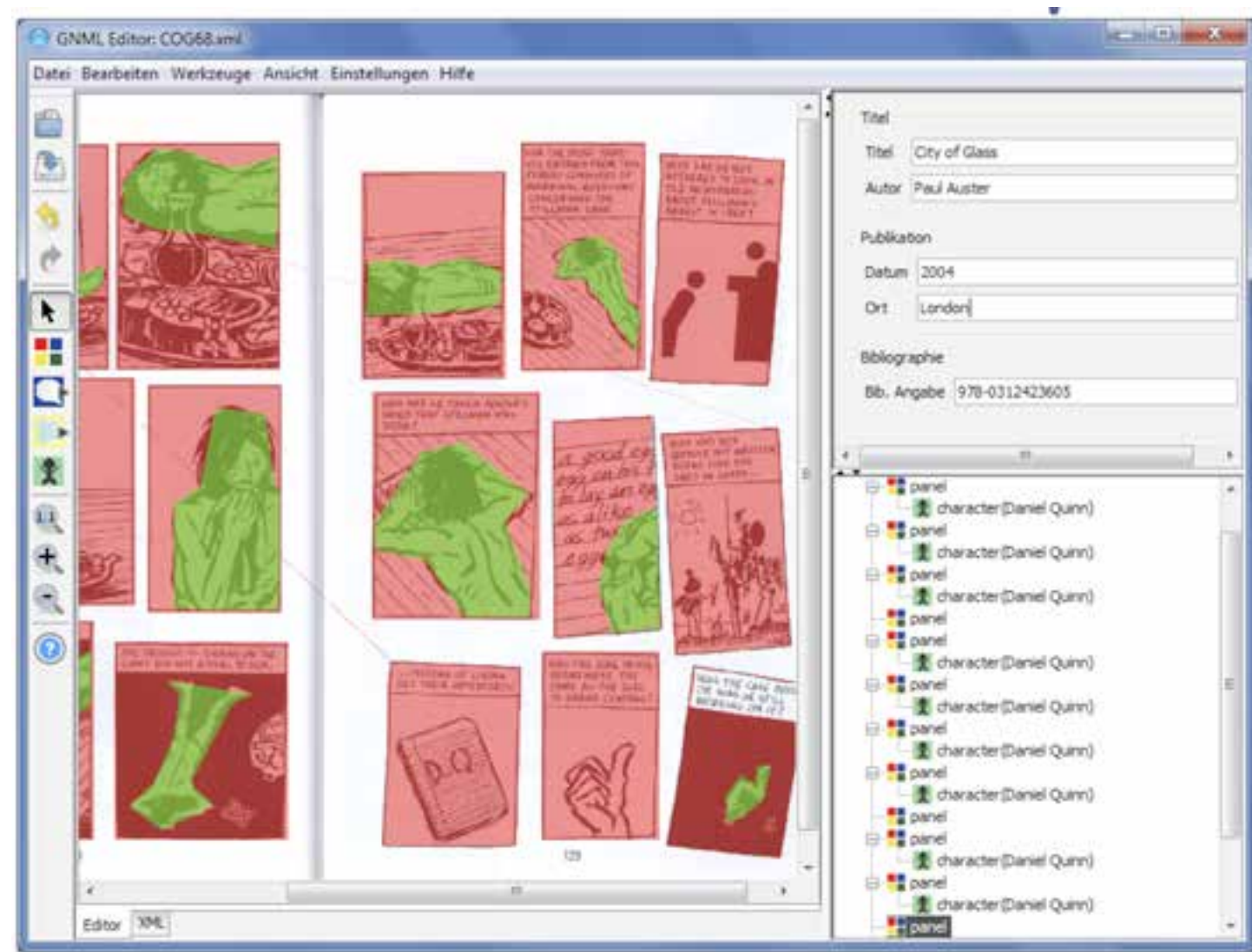


### Abstract

Whereas processes of reading and scene perception have both been researched extensively in cognitive psychology, little is known about their integration, as required by media with hybrid content such as graphic novels or comics strips, or illustrations in textbooks, manuals, or assembly instructions. How do text and image interact? What is the relative amount of time spent on decoding text vs. image? How effectively can overt attention select relevant parts of the current and upcoming scene? Can cognitive effort be related to different forms of panel transitions (McCloud, 1993)? Here we present first results from an ongoing investigation of eye movements during the reading of graphic novels and other forms of sequential art, eventually leading to the Potsdam Comics Corpus (PoCoCo) of eye movements on graphic novels. Data from the eye movement corpus will be enriched by experimental data. Results suggest that text in speech bubbles and captions is selected first, before aspects of the image are fixated. Much more time is spent on decoding the text than decoding the image. Within the image, characters are fixated more often than inanimate structures and there seems to be a clear guidance towards plot-relevant parts of the scene. Lower-level saliency also exerts an influence, but of a relatively minor role. Panel transitions can clearly be related to cognitive effort, suggesting that pre-processing of the upcoming panel is effectively performed by peripheral vision. Taken together, results suggest that text processing plays a relatively important role in graphic literature. Aspects of the visual scene are processed much faster, and parafoveal preview is effectively used to guide the gaze.

### Method



#### XML-based annotation language

GNML, extends CBML, a TEI XML scheme; AOs and tags for characters and objects, panel transitions

#### Graphical editor

for visual annotation, generates GNML, supports automatic detection of panels and further objects



#### Eye tracking

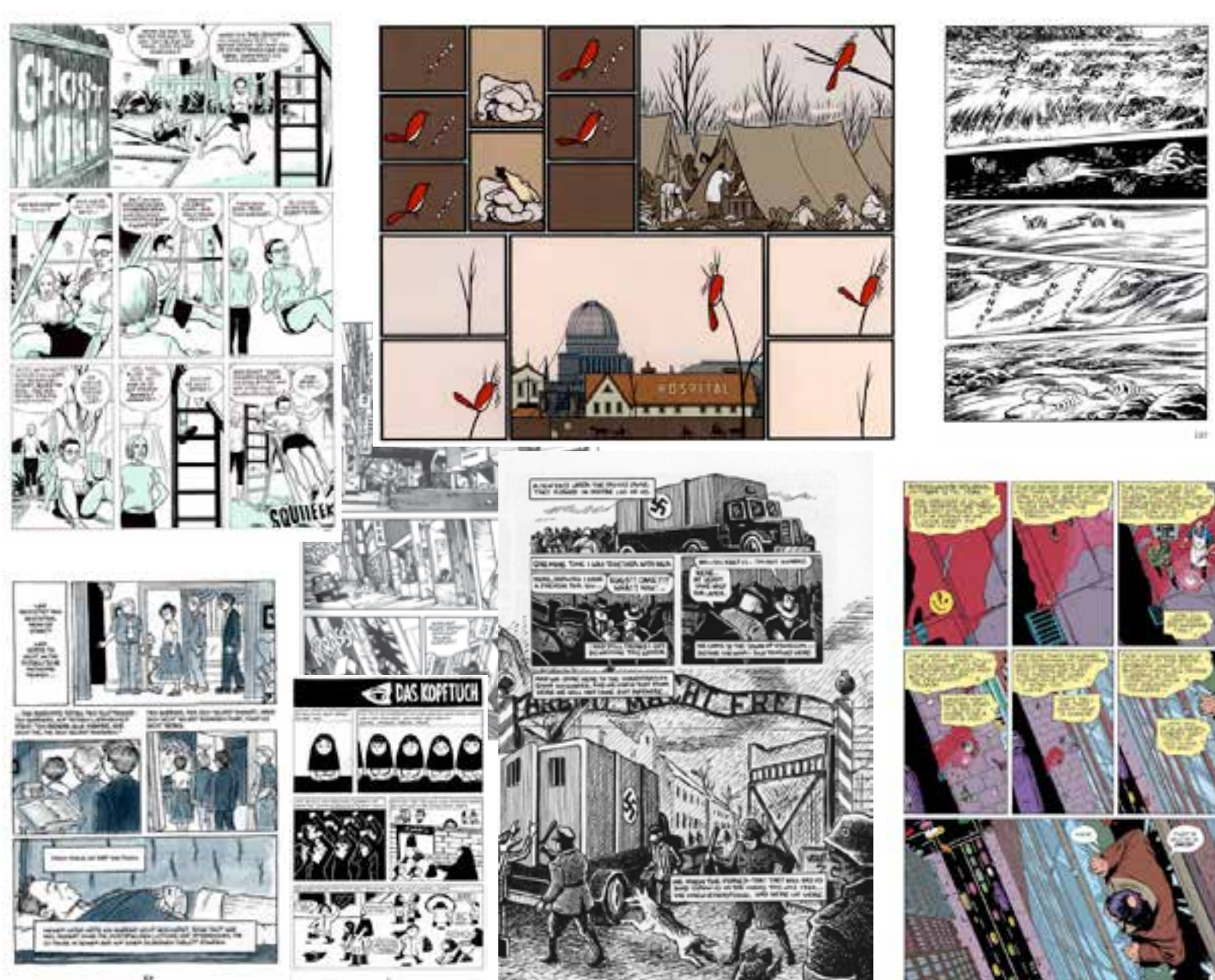
EyeLink 1000 (SR research), standard viewing (present results) plus gaze-contingent manipulations (future); high-resolution (5K) display to allow presentation of two pages with small

#### Reference corpus

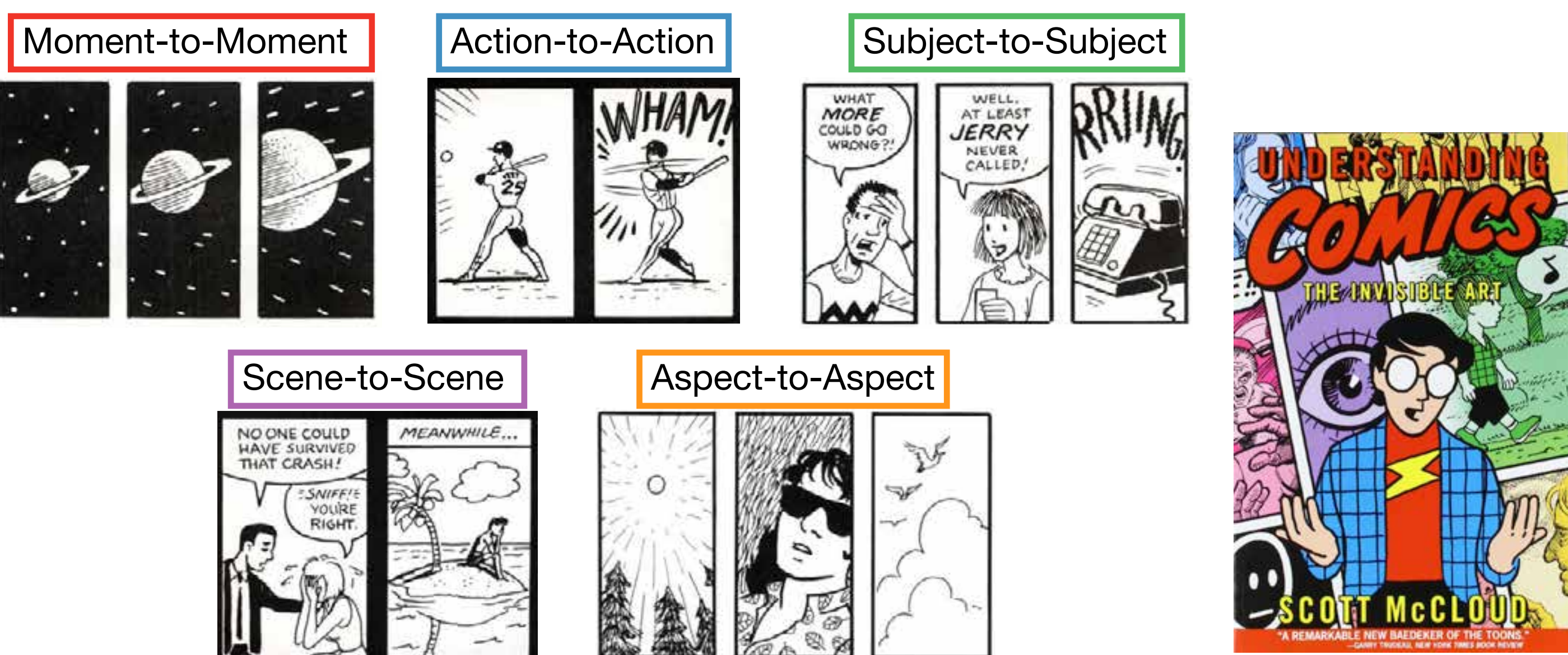
Canonical graphic novels plus annotations plus computer vision descriptors plus eye movement recordings (work in progress)

For comparison, inclusion of more standard comic books is planned

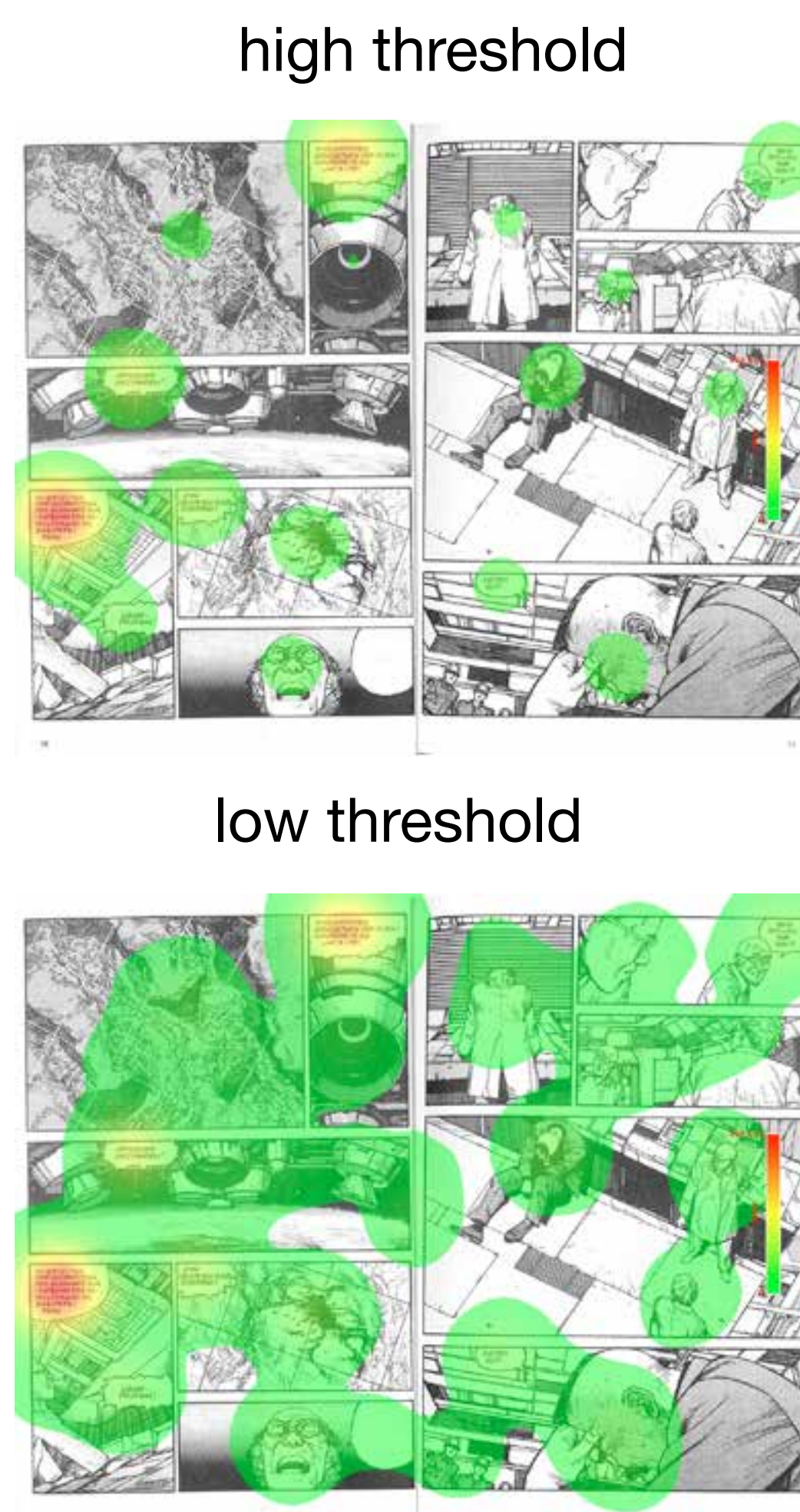
R packages for corpus and eye movements data analysis



### Experiment/case study: Test of some of Scott McCloud's hypotheses

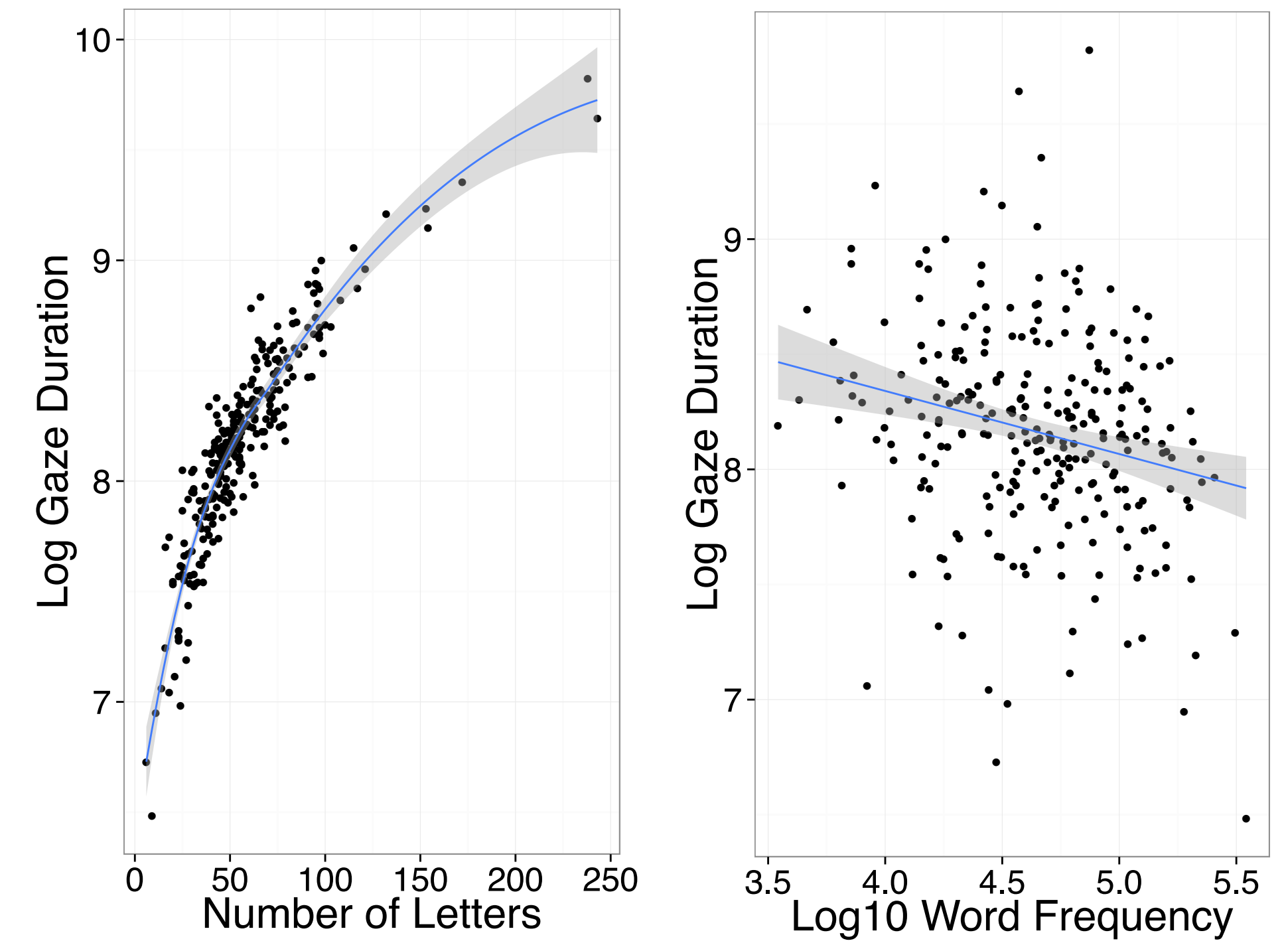


1. Classes of Panel transitions differ in the cognitive effort required from the reader
2. Time passes in the gutter



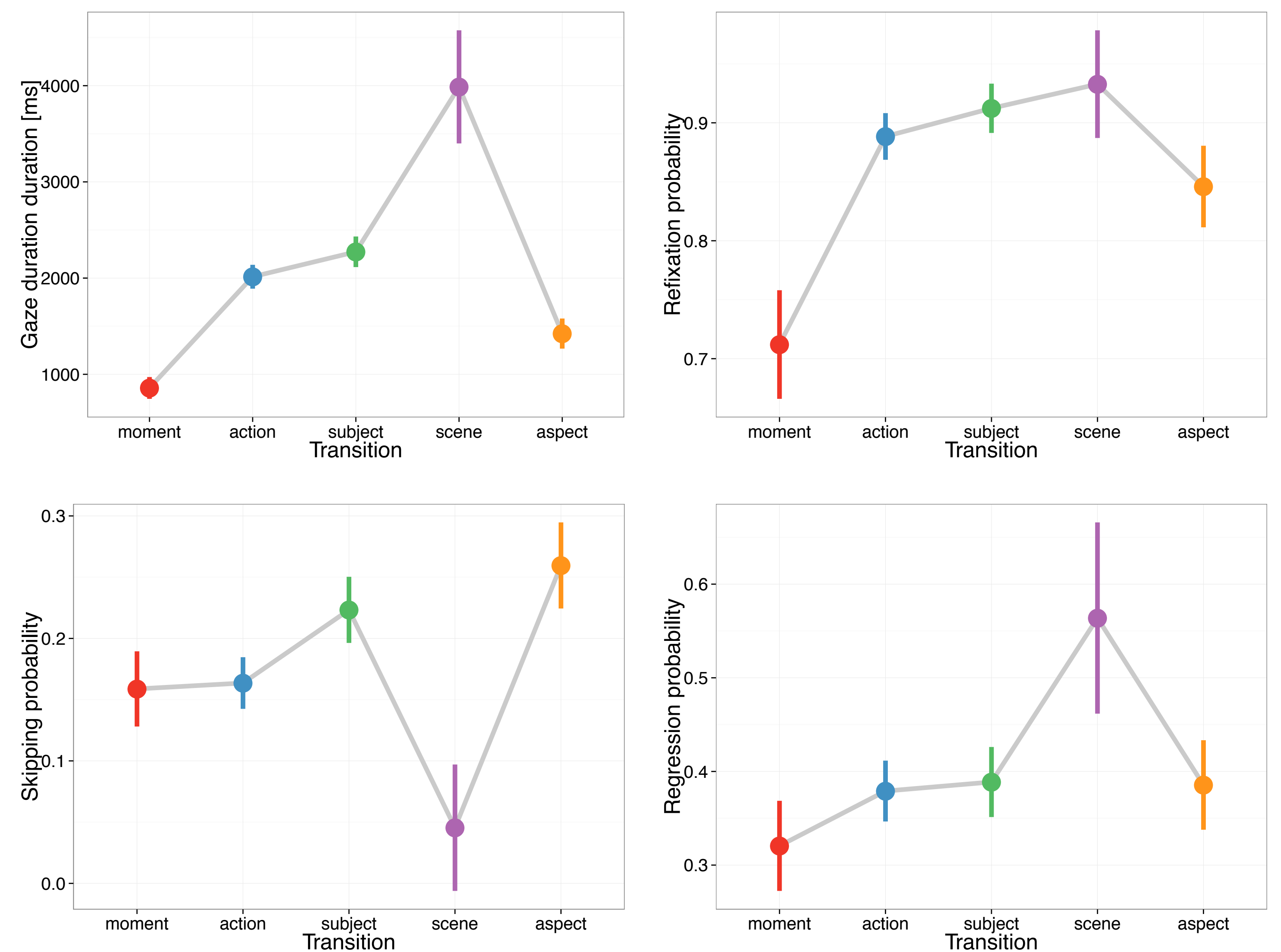
### Results

#### Distribution of Attention



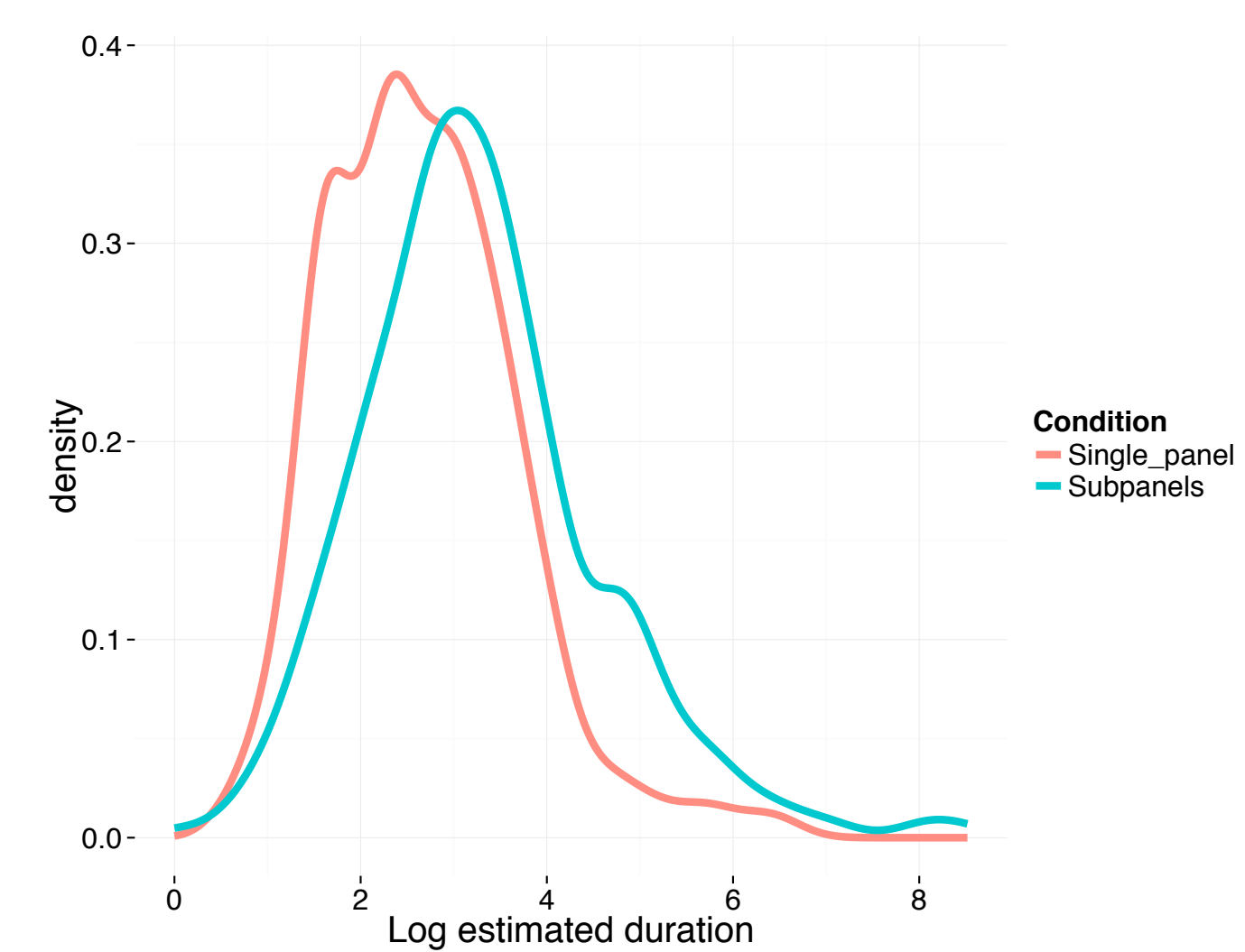
Attention is mostly attracted by text, faces and people. Other parts of the scene are also fixated—especially if they are referred to in the text. Low-level saliency does play a role, but a relatively minor one. Detailed and structured parts of the background do receive fixations; parts of the image with little structure, detail or texture hardly receive any, suggesting that fixations on the background mainly serve to analyze the gist of a scene, which can be analyzed quickly in a single fixation. Panel-averaged word frequency and cumulative length effects indicate cognitive processing of the text.

#### Panel Transitions



Eye movement measures support McCloud's typology. Moment transitions require the least effort, action and subject transitions require more effort, and scene transitions are the hardest to process, because they require inferences („closure“). Aspect transitions (establishing a mood, set a wandering eye on different aspects of a place) seem fairly easy to process, but are possibly just ignored by Western readers, since they do not immediately contribute to the storyline.

#### Time passes in the gutter



Subjective time estimates (“How much time has passed in this sequence?”, between Ss) were longer if a complex single panel was split into subpanels. This was true for each of the 10 tested panels. Thus McCloud's hypothesis that the gutter acts to indicate passing of time receives support.

### References

Cohn, N. (2013). *The visual language of comics: Introduction to the structure and cognition of sequential images*. London: Bloomsbury.  
McCloud, S. (1993). *Understanding comics*. Northampton, MA: Kitchen Sink Press.