By extending a recently published model of saccade generation (SceneWalk model; Engbert et al., 2015; Schütt et al., 2017) we were able to account for the empirical data. The model is based on two competing pathways that provide potential saccade targets (attention map) and keep track of recently fixated locations (fixation map). To generate a strong early CFB, we needed to assume that the sudden image onset led to a strong central activation in the attention map. The dynamic model was the only model to qualitatively reproduce the CFB and the relation between saccade latency and the distance to center of the second fixation. However, in its current form the model underestimated both effects. A control model that randomly selected saccade targets from the distribution of empirically observed fixation locations (Density model) performed worse than the SceneWalk model but demonstrated that fixations are distributed randomly and proportional to the empirical distribution on the image after about 2 s. Note, a very similar target selection mechanism is often assumed in saliency models where targets are sampled randomly from a saliency map. By incorporating systematic eye movement tendencies these models improve (c.f., Le Meur & Liu, 2015). A model similar to the SceneWalk model but without a fixation map (Gauss Model) performed well in terms of likelihood but did not reproduce the temporal evolution of the CFB or influences of saccade latencies. Finally, a pure central fixation bias model (CenterBias model) performed poorly on all measures.

Our results imply to use a modified version of the scene viewing paradigm to study bottom-up and top-down processes of target selection beyond the CFB. To minimize the influence of the sudden image onset, we suggest to use a fixation marker that disappears about 125 ms after image onset. In addition, due to the dependence of successive fixations, scene exploration should not exclusively start near the image center. Instead initial fixations (fixation markers) should be evenly distributed across the entire image or even with a preference towards the periphery. Central parts of the image will be fixated when the