We conducted data simulations to evaluate the power of the planned analyses. The following power analysis was conducted with a sample size of N=16. We assumed a mean proportion of looks into the mechanism of 0.15 or 0.25 for the functional condition and 0.4 or 0.5 for the non-functional condition. We simulated the random intercepts of subject ID by drawing random numbers (one per subject) from a normal distribution with a mean of zero and a standard deviation of 1.1. The size of the standard deviation corresponds to the coefficient for the intercept (qlogis(0.25)). Next, we simulated random slopes of condition within subject by drawing random numbers from a normal distribution with a mean of zero and a standard deviation of corresponding to the magnitude of the coefficient for the differences between conditions (qlogis(0.25)-qlogis(0.5); resulting in a random slope of 1.1). We simulated random slopes of session and trial number within subject by drawing random numbers from a normal distribution (one per subject) with a mean of zero and a standard deviation being 0.2. We then combined the fixed and random effects into the linear predictor and generated the response variable by sampling from a binomial distribution using the inverse logit transformed linear predictor as probability of success. We simulated 1000 datasets (with 16 subjects and 8 trials per condition) and fitted the GLMM described in the data analysis section (looking ~ condition + z.trial + z.session + (1 + condition.c + z.trial + z.session | subj.id)). We evaluated the models using two criteria: whether they (i) converged and (ii) whether the likelihood ratio test of condition was significant. We determined the proportion of models that fulfilled both criteria out of all simulated models. This revealed a power of 74% with performance levels of 0.15 in the functional condition and 0.4 in the non-functional condition and a power of 66% with performance levels of 0.25 in the functional condition and 0.5 in the non-functional condition.