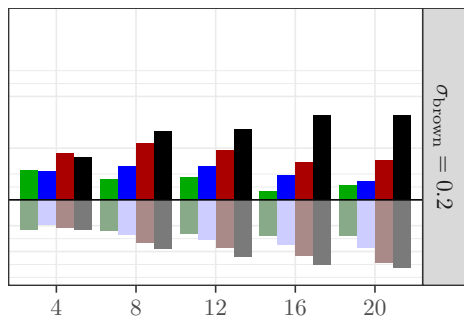
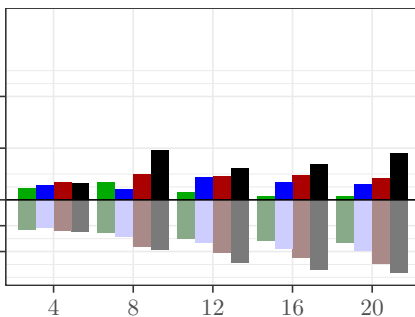
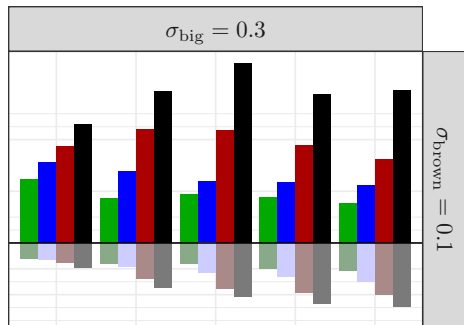
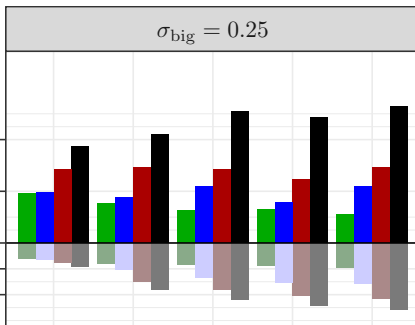


difference in mean success probability /
proportion of exceptions



number of objects in context

semantic threshold 0.2 0.4 0.6 0.8