Supplementary Material: Anticipation in Real-World Scenes: The Role of Visual Context and Visual Memory

Moreno I. Coco
Faculdade de Psicologia, Universidade de Lisboa
Alameda da Universidade, Lisboa 1649-013, Portugal
micoco@fp.ul.pt

Frank Keller
School of Informatics, University of Edinburgh
10 Crichton Street, Edinburgh EH8 9AB, UK
keller@inf.ed.ac.uk

George L. Malcolm
Department of Psychology, The George Washington University
2125 G Street NW, Suite 304, Washington, DC 20015, USA
glmalcolm@email.gwu.edu

Analysis of the empirical logit of fixation on the target region during the verb phrase when aggregated over time

In this section, we demonstrate that anticipatory looks triggered by verb specificity are also found when fixation data is aggregated across the whole linguistic region of interest.

Figure 1 plots the mean empirical logit of fixations aggregated over a 600 ms window (i.e., from 100 ms to 700 ms after the onset of the verb) for the four experimental conditions in Experiment 1 (left) and Experiment 2 (right). Consistently across the two experiments, we observe more anticipatory looks to the target region when the verb is specific compared to when it is ambiguous. Table 1 presents the results of fitting a mixed effects model on the aggregated data, showing a significant main effect of Verb for both experiments.

This finding is consistent with the results of the analysis using time bins reported in the main paper, where the anticipation effect manifested itself as an interaction Verb:Time. The advantage of that analysis is that it makes it possible to track the time course of fixations across the window of analysis, rather than just providing an aggregate measure of fixations over the whole region.

Empirical logit of fixation at the target region across the whole sentence

In this section, we report how attention to the target region distributes over the course of the sentence, and test whether at the onset of the verb phrase, which is our linguistic region of interest, fixations significantly differ between experimental conditions. If anticipatory looks to the target region are genuinely triggered by the processing of verb specific information, we should observe no significant difference at the verb onset.

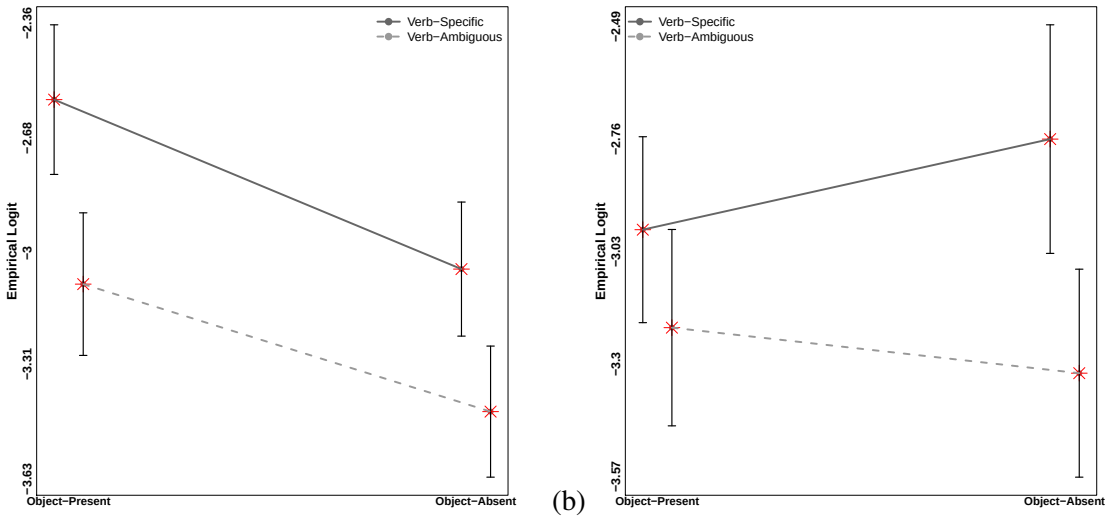


Figure 1. Interaction plot of the empirical logit of fixation (mean and standard error) on the target region during the verb phrase, aggregated over the linguistic region of interest for the experimental conditions of Object (Present, Absent) and Verb (Specific, Ambiguous): (a) Experiment 1, (b) Experiment 2. The asterisk represents the mean of the fitted mixed effects model.

(a) *Experiment 1*

Predictor	β	*SE*	*t*	*p*
Intercept	-3.04	0.16	-18.86	<0.0001
Verb	0.44	0.18	2.38	0.01
Object	0.44	0.26	1.68	0.1
Object:Verb	0.13	0.36	0.37	0.7

(b) *Experiment 2*

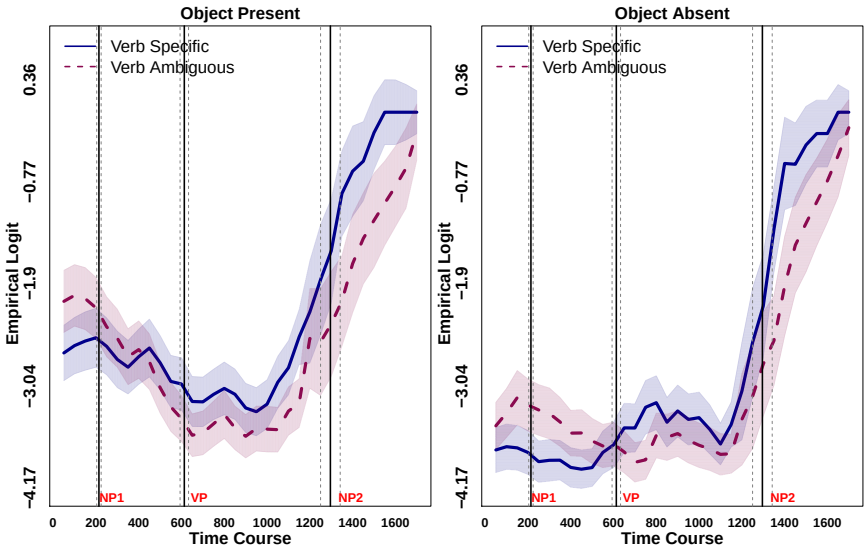
Predictor	β	*SE*	*t*	*p*
Intercept)	-3.12	0.31	-9.9	<0.0001
Verb	0.4	0.2	1.95	0.05
Object	0.02	0.21	0.13	0.8
Object:Verb	-0.33	0.41	-0.8	0.4

Table 1

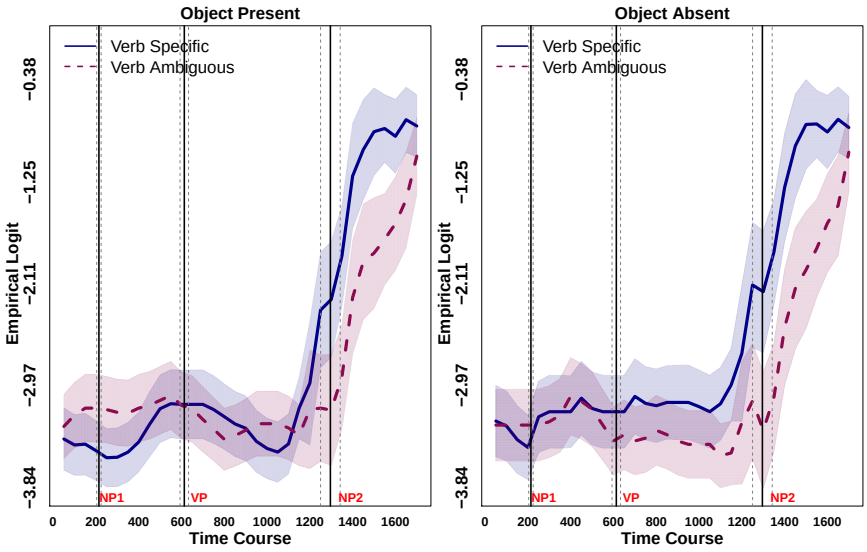
Aggregated fixations: Results of the mixed effects model with maximal random effects structure. The dependent measure is empirical logit of the fixation probability on the target region aggregated over a window of 600 ms aligned at the verb onset. The predictors are Verb (Specific, 0.5; Ambiguous,−0.5) and Object (Present, 0.5; Absent, −0.5). Random slopes of Participants and Items and random intercepts for main effects were included in the model.

In Figure 2, we plot the empirical logit of fixations over the course of the sentence across experimental conditions (Verb and Object), for both experiments. The plots show that at the onset of the sentence, fixations on the target region are similarly distributed across experimental conditions. Only after the verb is processed, we observe a rise in the fixation to the target region. This effect is consistent for both experiments. These results are consistent with what we report in the main paper.

Moreover, in order to test that there was no significant difference across experimental conditions at the verb onset, we calculated the empirical logit of fixation within a window of 400 ms, centered at the verb onset (i.e., verb onset ±200 ms), and conducted an additional mixed effects



Experiment 1



Experiment 2

Figure 2. Time-course plots of the empirical logit of fixation on the target region from beginning to end of the sentence (0 ms to 1700 ms) centered at the onset of the verb for the different experimental conditions for the two Experiment (top row, Experiment 1; bottom-row Experiment 2). For each experiment, we group the experimental condition in Object Present (left panel) and Object Absent (right panel). The shaded bands indicate the standard error around the observed mean, which is plotted as a line. On each plot, we mark the mean onset of the linguistic regions of interest (NP1 *the man*, VP *ate*, NP2 *the sandwich* in our example), as well as their confidence intervals.

(a) *Experiment 1*

Predictor	β	*SE*	*t*	*p*
Intercept	-3.48	0.15	-21.84	<0.0001
Object	0.42	0.25	1.66	0.1
Verb	0.24	0.23	1.04	0.2
Object:Verb	0.22	0.46	0.47	0.6

(b) *Experiment 2*

Predictor	β	*SE*	*t*	*p*
Intercept	-3.26	0.35	-9.11	<0.0001
Object	0.23	0.32	0.70	0.4
Verb	0.12	0.24	0.53	0.6
Object:Verb	-0.23	0.48	-0.47	0.6

Table 2
Verb onset: Results of the mixed effects model with maximal random effects structure. The dependent measure is empirical logit of the fixation probability on the target region calculated within a window of 400 ms centered at the verb onset. The predictors are Verb (Specific, 0.5; Ambiguous, −0.5) and Object (Present, 0.5; Absent, −0.5). Random slopes of Participants and Items and random intercepts for main effects were included in the model.

model analysis with Verb and Object as predictors, and Participants and Items as random slopes and intercepts. The model shows no significant main effect of Verb and Object, and no significant interaction (see Table 2 for the model coefficients).

Correlation between fixations on target region and number of objects in the scene

This section shows that the attention to the target region during the processing of the linguistic region of interest is not associated with the number of visual object present in the scene.

We calculated the empirical logit of fixation on the target region, aggregated over a 600 ms window (i.e., from 100 ms to 700 ms after the onset of the verb, same as in the first analysis in the Supplementary Material). A Pearson product-moment correlation coefficient was computed to assess the relationship between the empirical logit of fixation on the target region in this time window and the number of objects in the scene. There was no significant correlation between these two variables either in Experiment 1 ($r = -0.05$, $n = 576$, $p = 0.25$), or in Experiment 2 ($r = -0.02$, $n = 576$, $p = 0.6$).

This analysis serves to demonstrate that differences between real-world scenes and clip-art scenes cannot be wholly driven by the increased number of recognizable visual objects that real-world scenes offer. Rather the effect should be attributed to the fact that real-world scenes provide context information, which can be used to guide the allocation of visual attention, making it possible to restrict attention to the objects relevant to the task (see main article for a more detailed argumentation to this effect).