Supplemental materials for "When the eyes have it and when not: How different sources of activation combine to guide eye movements during multi-attribute decision making"

Agnes Rosner, Michael Schaffner, Bettina von Helversen

Feedback Procedure During Exemplar Learning in Experiment 1

When participants made a correct response, the attribute value appeared in the rectangle. The color of the rectangle changed from white to green and a high-pitched tone was presented to indicate a correct response. Feedback was visible for 3 s. If a wrong response was made, for 1 s the value of the rectangle the participant had chosen appeared in red. The rectangle turned red and a low-pitched tone was presented to indicate a wrong response. Immediately after that, the wrong attribute value disappeared and the correct value appeared in the correct location for 2 s. A color change from white to green of the correct rectangle indicated that this was the correct location for the auditorily presented word.

Feedback Procedure During Criterion Learning in Experiment 1

First, the color of the outline of a person in the background of each candidate (see Figure 2 of the main text) changed from white to green if the candidate was invited and to red if the candidate was rejected. Second, a statement in the center of the screen indicated whether the choice was correct and whether this candidate was invited. Last, a tone was presented indicating if the choice was correct or not. For instance, in the case of correctly rejecting a candidate, the color of the person's outline changed from black to red and a statement appeared: "Correct! The candidate was rejected" and a high-pitched tone was presented auditorily.

LAN Over Time in Experiment 1

The high number of participants showing LAN behavior in the Auditory condition may have been due to participants having already started looking at the empty spatial locations during the auditory presentation of the test candidates, reflecting language

processing (Huettig et al., 2011). In contrast, participants in the visual conditions first had to read the verbal description of the candidates presented in the center of the screen before they could begin LAN. To test this hypothesis, we analyzed LAN over time. We determined mean proportions of fixation on the exemplar locations per 250-ms time bin. Figure S1 shows mean fixation proportions over all trials and participants for all experimental conditions. In conditions with the removal of test stimulus presentation (Auditory, 3-s Visual, 1.5-s Visual), fixation proportions after the removal were analyzed for time bins normalized on trial response times.

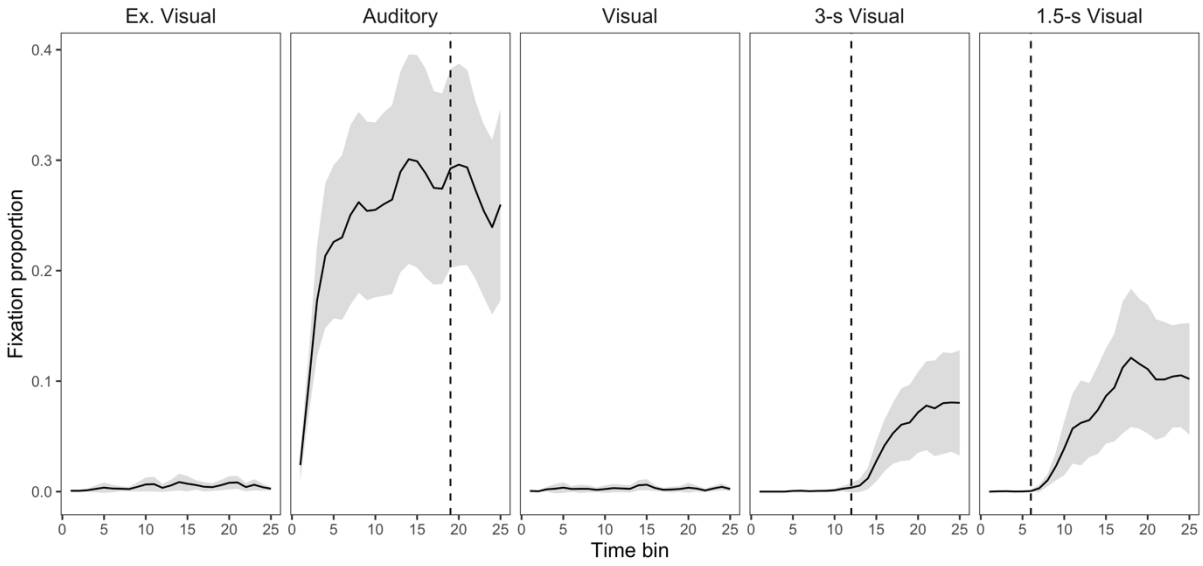

Figure S1. Mean proportions of fixation per time bin on the four exemplar locations. Dashed lines indicate the end of the third auditorily presented attribute in the Auditory condition and the removal of the test item from the center of the screen in the 3-s and 1.5-s Visual conditions. Bin size was 250 ms. Bin sizes after the removal are normalized on response times. Gray shaded ribbons show 95% confidence intervals around the mean.

LAN was most pronounced in the Auditory condition and was not observed on the aggregate level in the Ex. Visual and Visual conditions, where information stayed visible on the screen. When comparing average mean onset times of the first fixation landing on one of

the four exemplar locations in cases where people showed LAN, we found that in the Auditory condition, LAN to exemplar locations had already begun during the onset of the auditory stimulus presentation (M_{Auditory} = 1.48 s, SD_{Auditory} = 0.80). LAN in the 1.5-s Visual condition began after the end of the visual presentation of the test item ($M_{\text{1.5-s Visual}}$ = 3.35 s, $SD_{\text{1.5-s Visual}}$ = 0.92). To test for this observed difference, we ran a linear mixed-model analysis, with time until the first look toward an exemplar as the dependent variable and condition (Auditory and 1.5-s Visual) as a fixed effect. We added by-subject random intercepts. The analyses revealed significant differences between onsets of LAN, $F(1,34.86) = 76.22$, $p < .001$.

Additional Results on LAN and Categorization Decisions in Experiment 1

None of the covariates significantly predicted categorization decisions: location memory performance, $\chi^2(1) = 1.62$, $p = .20$; response time, $\chi^2(1) = 1.07$, $p = .30$; interaction response time with condition, $\chi^2(4) = 7.58$, $p = .11$.

Detailed Analyses of Categorization and Memory Performance in Experiment 2

Categorization Accuracy

To test how well participants performed in the categorization task, we analyzed responses on the exemplar items that could be clearly classified. Categorization accuracy was lower than in Experiment 1, indicating that the task was more difficult. For statistical comparison, we ran a mixed-model analysis with a fixed effect for the instruction and by-subject random intercepts and by-subject random slopes for items (i.e., the six exemplars). We did not observe meaningful differences between the groups in categorization accuracy, $\chi^2(1) = 0.48$, $p = .49$.

Categorization Response Times

Categorization response times were measured from the onset of the screen containing attribute information until participants' response. For the comparison of categorization

response times, we analyzed data of all item types (exemplars, ambiguous items). We included fixed effects for instruction, item type (1 = exemplar, 2 = ambiguous item), and their interaction. As items were tested in blocks (ranging from 1 to 10), we included block as an additional fixed effect to test if participants got faster over the course of the experiment. As random effects, we included by-subject intercepts and by-subject random slopes for items. There was no difference between the instruction conditions, $F(1, 58.03) = 0.37$, $p = .55$, or between item types, $F(1, 66.81) = 1.92$, $p = .17$, and no interaction between instruction and item type, $F(1, 66.22) = 0.21$, $p = .65$. However, there was a main effect of block. Block determined categorization response times, $F(1, 5722.10) = 659.43$, $p < .001$. That is, participants got faster over the course of the experiment (linear contrast: $z = -21.68$, $p < .001$).

Location Memory Test

As an indicator of location accuracy, we summarized location memory pre- and posttest performance into one location memory score. The score indicates the number of correct localizations, with 2 indicating an exemplar that was correctly localized in the pre- and posttest, 1 indicating an exemplar that was correctly localized in either the pre- or posttest, and 0 indicating an exemplar that was never localized correctly. Thus, the higher the score, the better participants remembered the exemplar locations. We compared the location memory scores of the intuitive ($M_{Int} = 0.7$, $SD_{Int} = 0.48$) and explicit ($M_{Exp} = 0.7$, $SD_{Exp} = 0.31$) conditions with a mixed model, including a fixed effect for instruction and by-subject random intercepts, and by-subject random slopes for exemplars. Location memory performance did not vary between the instruction conditions, $F(1, 58.01) = 0$, $p = .96$.

Old–New Discrimination Test

Last, we examined if participants in the two instruction conditions varied in their ability to differentiate the exemplar items from new items. We compared discrimination accuracy between conditions with a mixed model, including a fixed effect for instruction and

by-subject random intercepts, and by-subject random slopes for exemplars. Discrimination accuracy did not vary between the instruction conditions, $F(1, 57.98) = 2.88$, $p = .10$.

Detailed Analyses of LAN Occurrence in Experiment 2

To test the effect of instruction on the likelihood of looking at the empty exemplar locations, we added instruction and item type as well as their interaction as fixed effects to a mixed-model analysis. Furthermore, we added block, response time, and categorization accuracy as covariates, and by-subject random intercepts. There were no differences in the likelihood of showing LAN behavior between the intuitive and explicit instruction conditions, $\chi^2(1) = 0.11$, $p = .74$. In addition, there were no effects of item type, nor an interaction between item type and condition, nor an influence of categorization accuracy (all ps > .56). There was only a main effect of categorization response time: $\chi^2(1) = 679.39$, $p < .001$ (linear contrast: $z = 22.52$, $p < .001$). As in Experiment 1, this indicates that with more time, participants became more likely to look at one of the exemplar areas of interest.

Furthermore, we found a main effect of block, with LAN becoming less likely over the course of the experiment, $\chi^2(1) = 5.02$, $p = .03$ (linear contrast: $z = -2.26$, $p = .02$). However, when running the same analysis only for the exemplar items and adding location memory performance and discrimination accuracy as further covariates, no effect of block was observed, $\chi^2(1) = 1.58$, $p = .21$. Other than an effect of response time, $\chi^2(1) = 419.27$, $p < .001$ (linear contrast: $z = 17.63$, $p < .001$), again, no other variable predicted LAN occurrence (all ps > .09).

LAN Over Time in Experiment 2

Following the results of Experiment 1, we expected that LAN would start occurring after the removal of the visual test stimulus and independent of the instruction condition. Indeed, in both instruction conditions, participants started with LAN after the removal of the visual test stimulus ($M_{Int} = 1.23$, $SD_{Int} = 0.66$; $M_{Exp} = 1.26$, $SD_{Exp} = 0.93$; Figure S2). The

same mixed model as for LAN occurrence, but with time until the first look at an exemplar as dependent variable, showed no meaningful difference between the instruction conditions, $F(1,37.40) = 0.06$, $p = .81$.

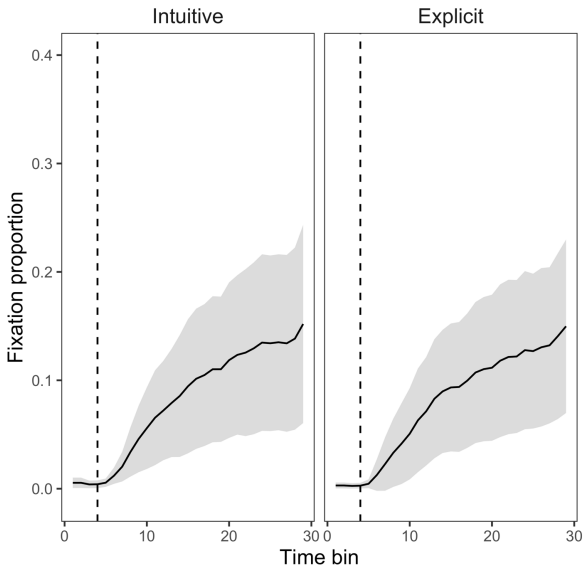

Figure S2. Mean fixation proportions per time bin for the six exemplar locations in the two instruction conditions of Experiment 2. Dashed lines indicate the removal of the test item from the center of the screen after 200 ms. Each bin before the removal had a size of 50 ms. Bin sizes after the removal are normalized on response times. Gray shaded ribbons show 95% confidence intervals around the mean.

Additional Results of the Analysis of LAN Strength in Experiment 2

There was an additional main effect of block. That is, LAN strength became weaker over the course of the experiment, $\chi^2(1) = 6.72$, $p = .01$ (linear contrast block: $z = -3.07$, $p = .002$).

Additional Results of the Analysis of LAN an Categorization Decisions

In addition, the longer participants took to respond, the more likely they were to decide on Category A, main effect of response time: $\chi^2(1) = 4.32$, $p = .04$. Although category

labels were reversed for half the participants, if they took longer to decide, they were more likely to press the left mouse button always indicating Category A.