Supplementary Figures

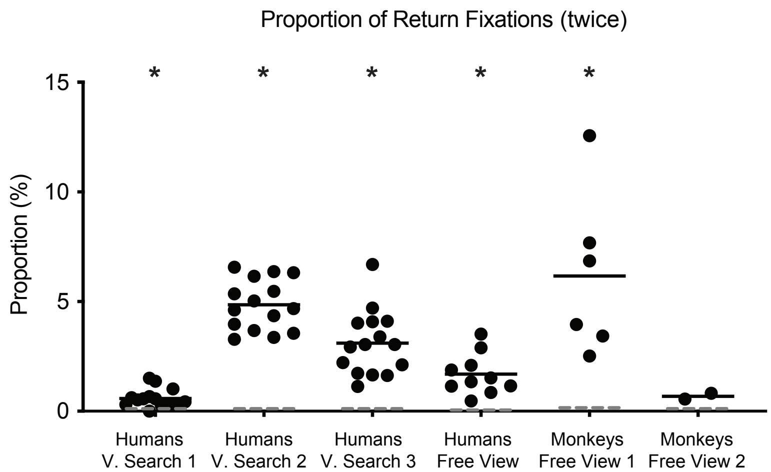


Fig S1. Proportion of fixations that revisit the same location twice. Following the format in **Fig. 3**, here we show the proportion of fixations that return to the same location twice (as in the sequence L_1-L_2-L_1-L_3-L_4-L_1 where L indicates different image locations). The black dots denote individual subjects (horizontal spread added for visualization purposes only). Solid lines show average across subjects. Asterisks indicate statistical significance with respect to chance. ($p < 0.05$, one-sample t-test).

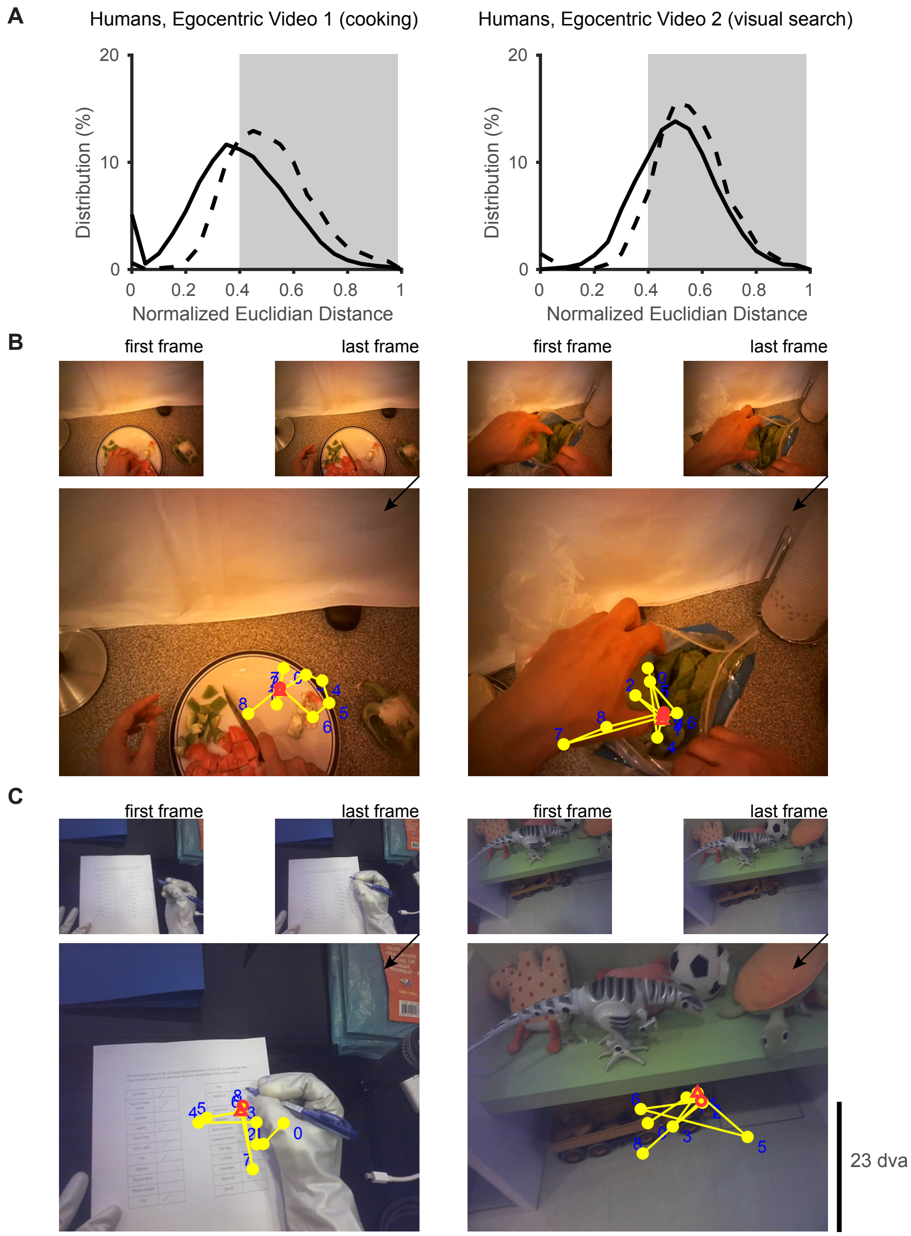


Fig S2. Extraction of fixations on egocentric videos. Egocentric videos (**Fig. 2G-H**) were analyzed in 5-second segments. (**A**). Distribution of normalized Euclidian distance between the first frame and the last frame (solid lines). The chance distribution of normalized Euclidian distance values was computed by considering random pairs of frames from 100 different video sequences (dashed line). Segments with large head motion (normalized Euclidian distance > 0.4, gray rectangle) were excluded from analyses. (**B-C**). Examples of the first and last frame in 5-second clips (top) and extracted fixations (yellow circles) mapped to the last frame of example video clips (bottom) during the cooking task (**B**, **Fig. 2G**) and visual search (**C**, **Fig. 2H**). The figure conventions follow **Fig. 1**.

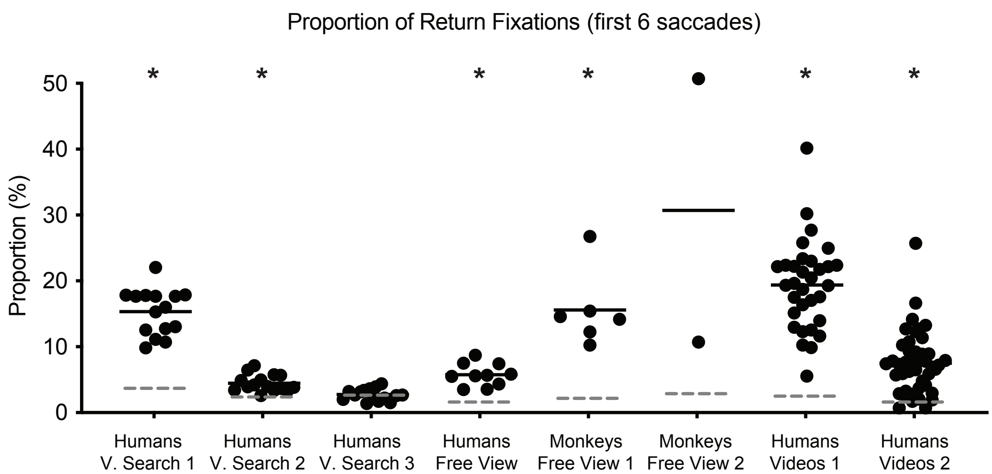


Fig S3. Proportion of return fixations among the first six fixations.
Following the format in **Fig 3**, here we show the proportion of return fixations among the first six fixations of all scanpaths over all eight datasets. Asterisks indicate statistical significance with respect to chance ($p < 0.05$, one-sample t-test).

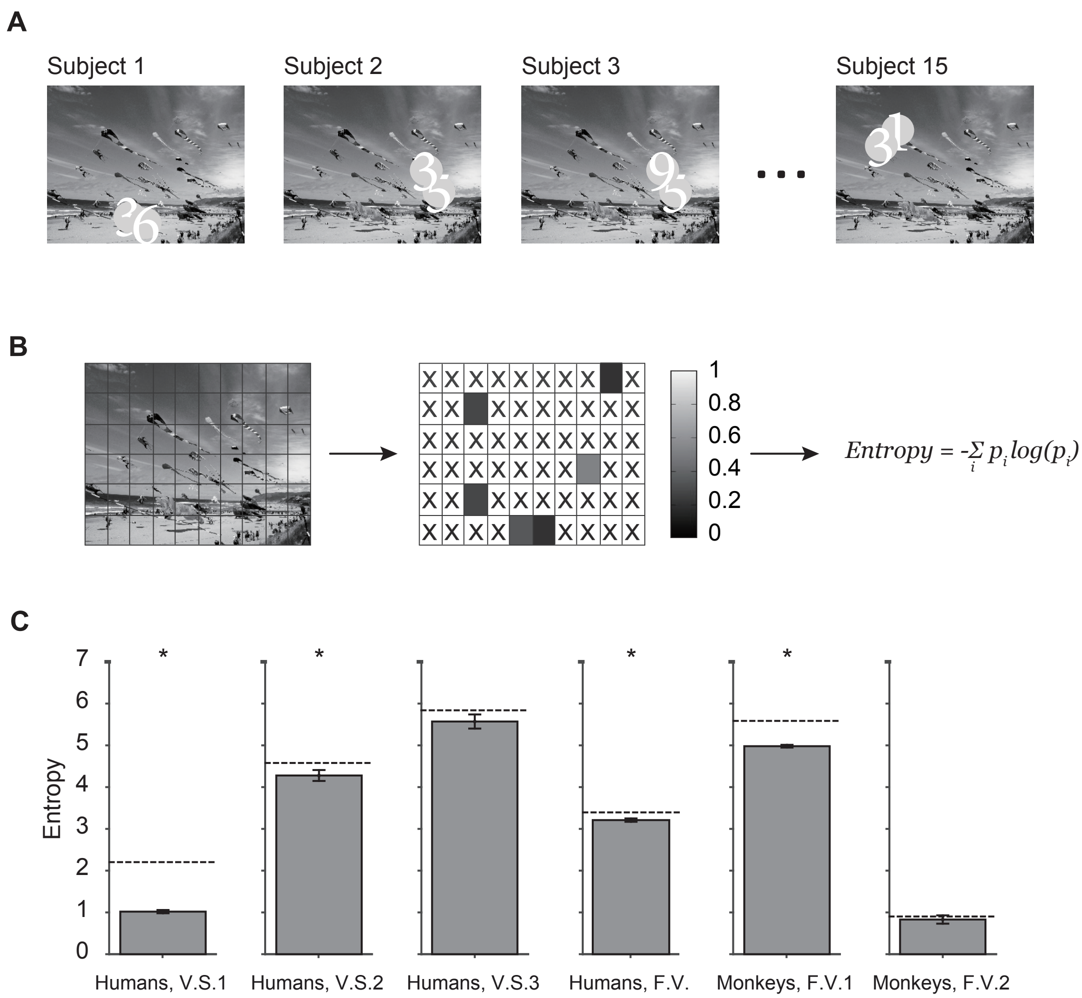


Fig S4. Return fixations are consistent across subjects. (**A**). Example return fixations over multiple subjects. The numbers indicate the fixation number in the sequence for each subject. (**B**). The image is divided into a 32×40 grid (only a few of the lines are shown in the image). For those locations where there is at least one return fixation, we compute the proportion of the total number of return fixations that land on that location (shown here by the grayscale colormap). The entropy of the distribution is then computed from those probability values. The lower the entropy, the higher the consistency between subejcts. (**C**). Entropy for each experiment. Horizontal dashed lines show the expected chance value obtained by assigning the total number of return fixations to random locations for each image.

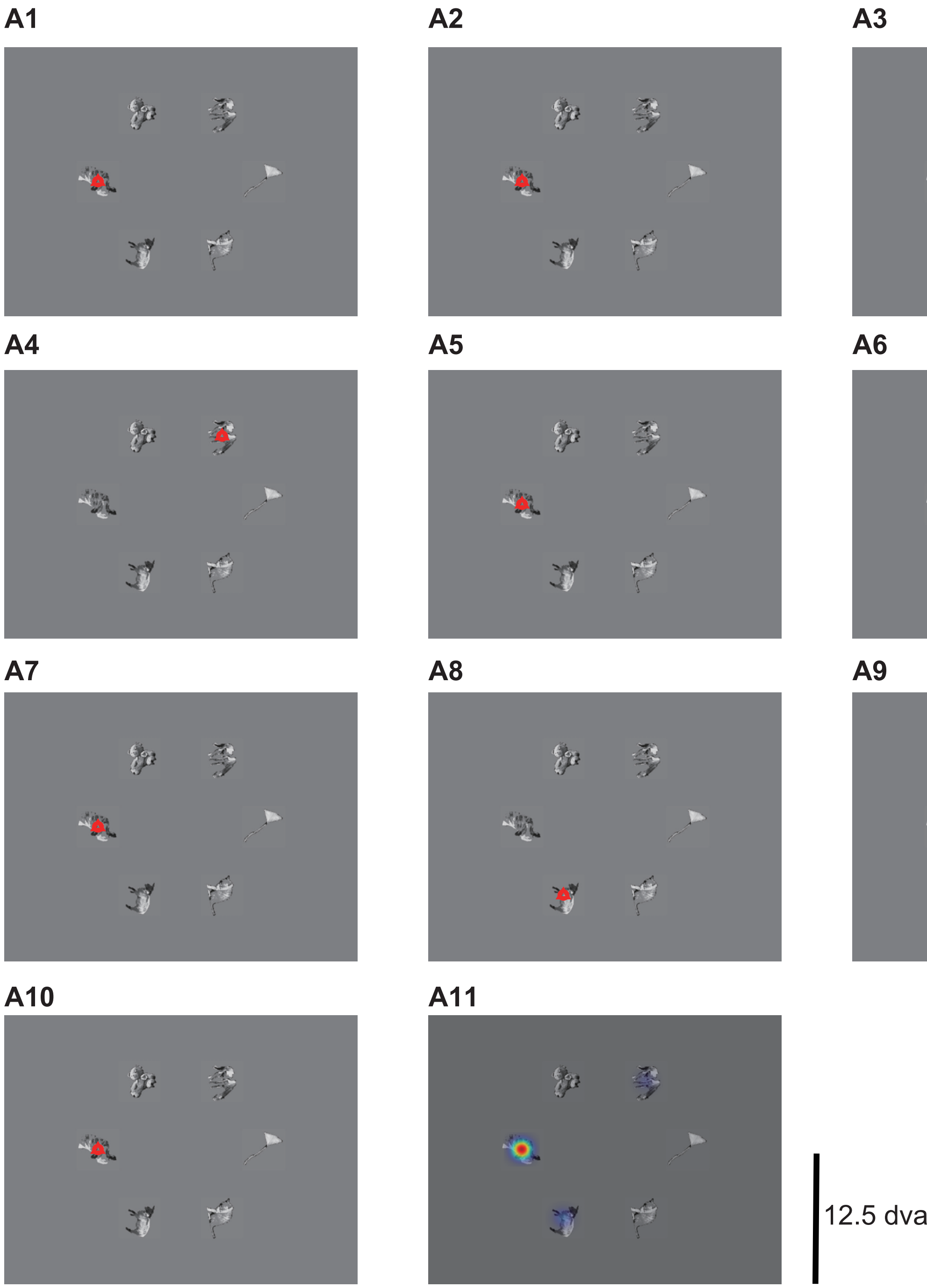


Fig S5. A. Example of consistent return fixations across subjects, Visual Search 1.

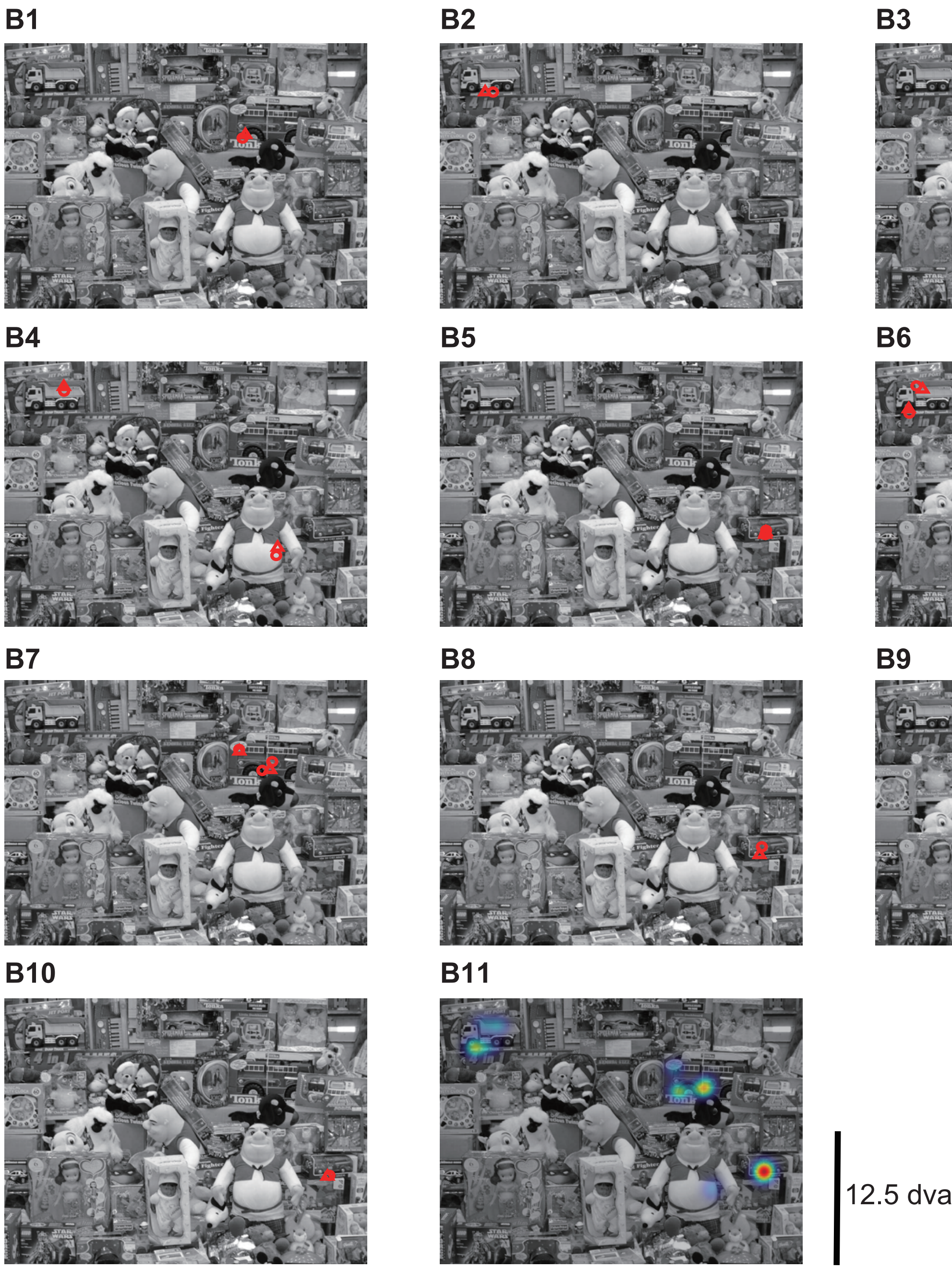


Fig S5. B. Example of consistent return fixations across subjects, Visual Search 2.

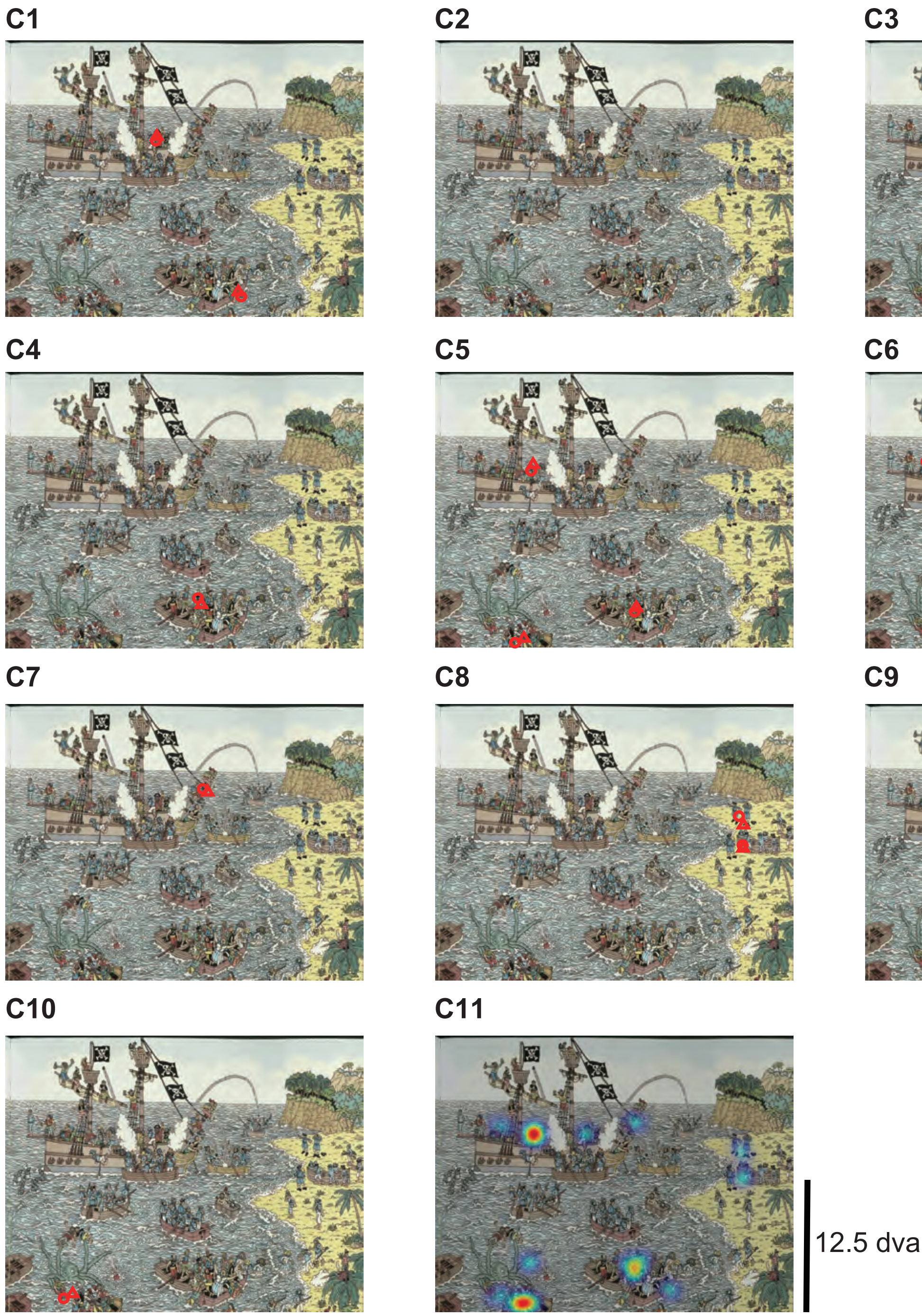


Fig S5. C. Example of consistent return fixations across subjects, visual search 3.

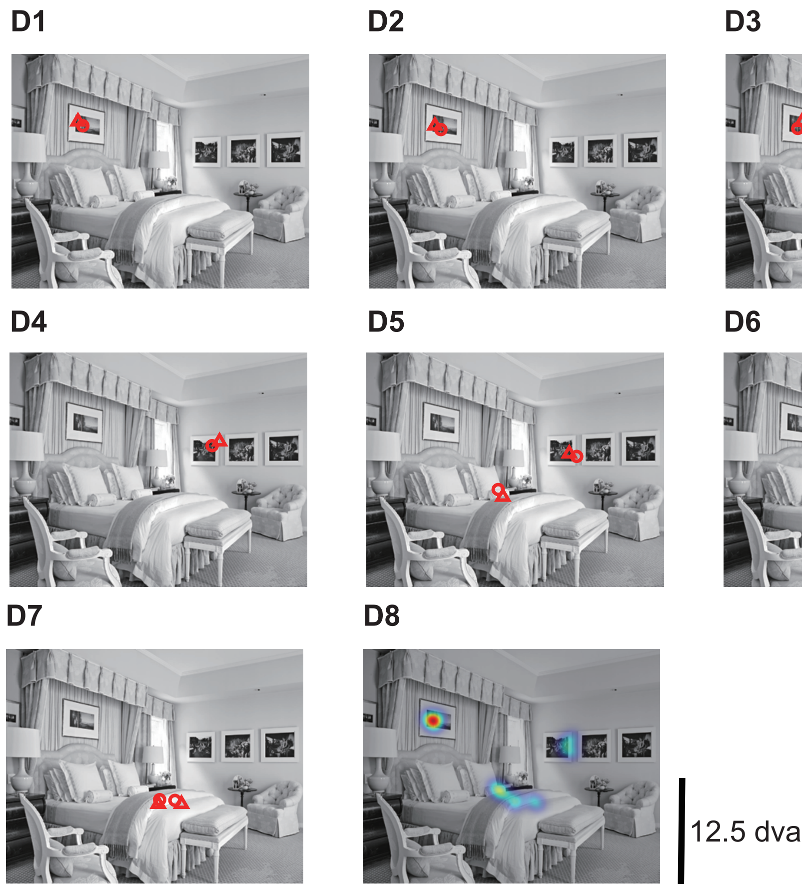


Fig S5. D. Example of consistent return fixations across subjects, free viewing (humans).

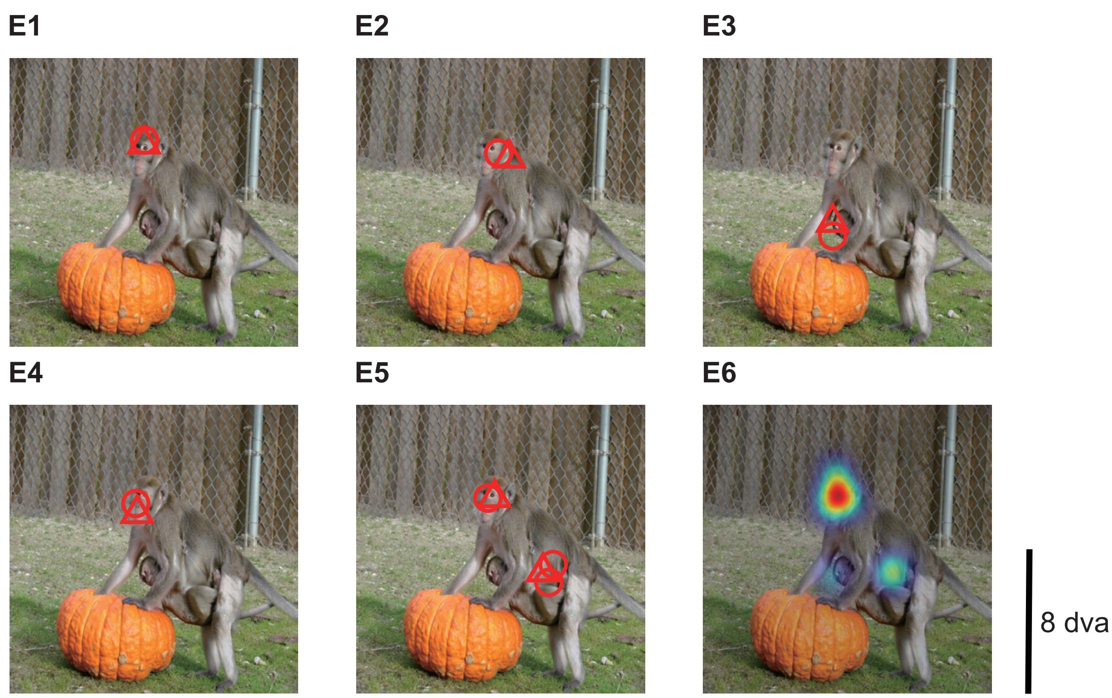


Fig S5. E. Example of consistent return fixations across subjects, free viewing 1 (monkeys).

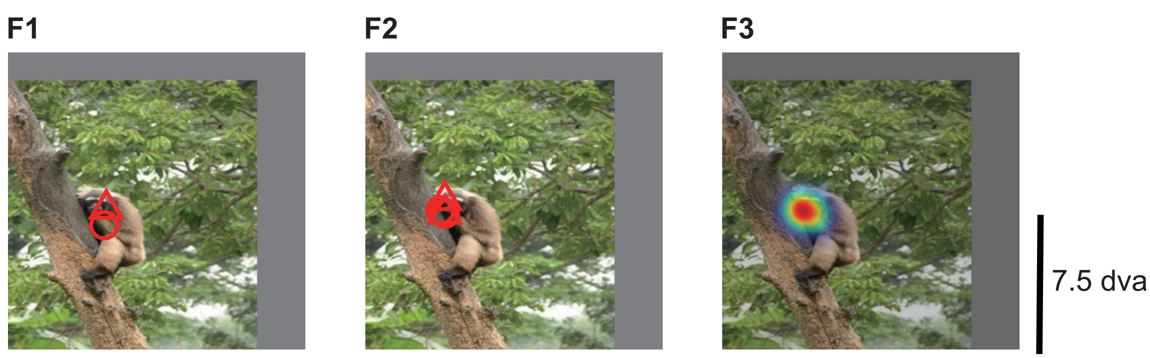


Fig S5. F. Example of consistent return fixations across subjects, free viewing 2 (monkeys).

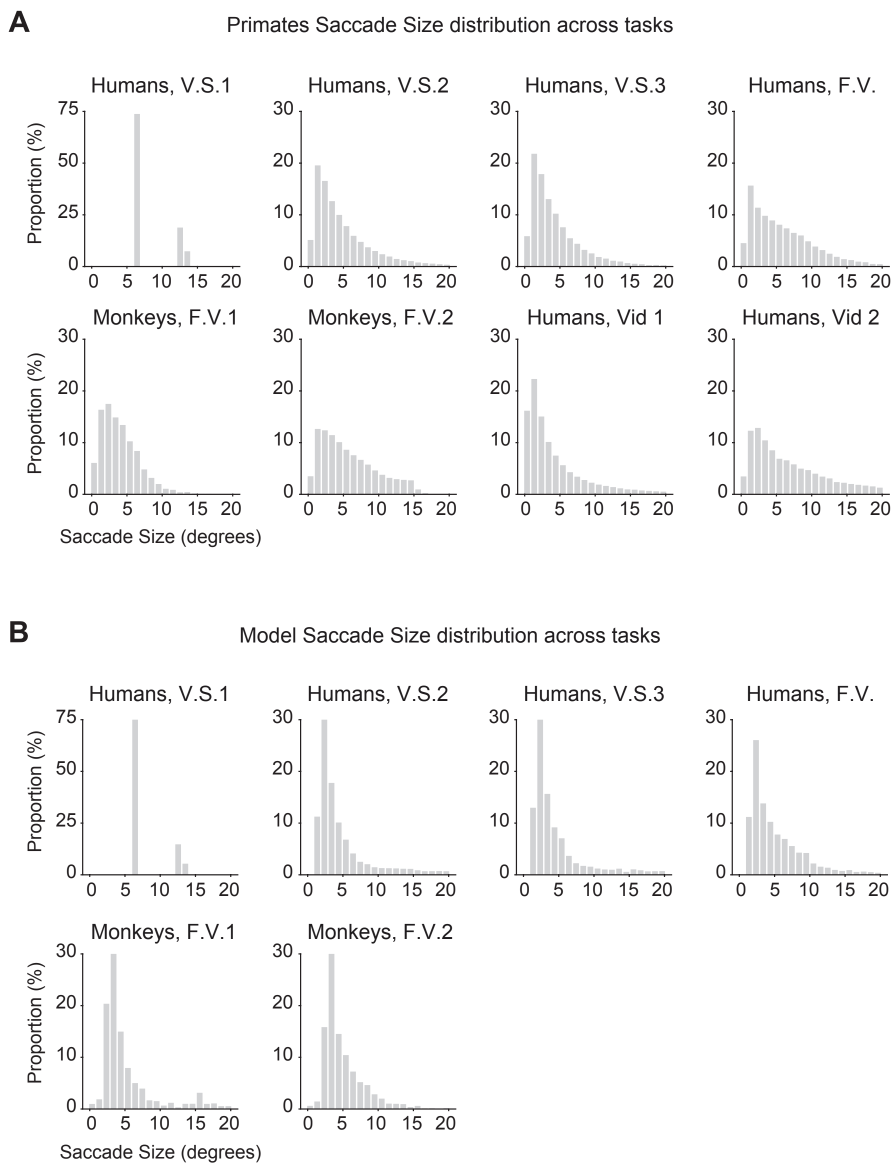


Fig S6. Distribution of saccade sizes. **A**. Distribution of saccade sizes for humans and monkeys. **B**. Distribution of saccade sizes for the model.

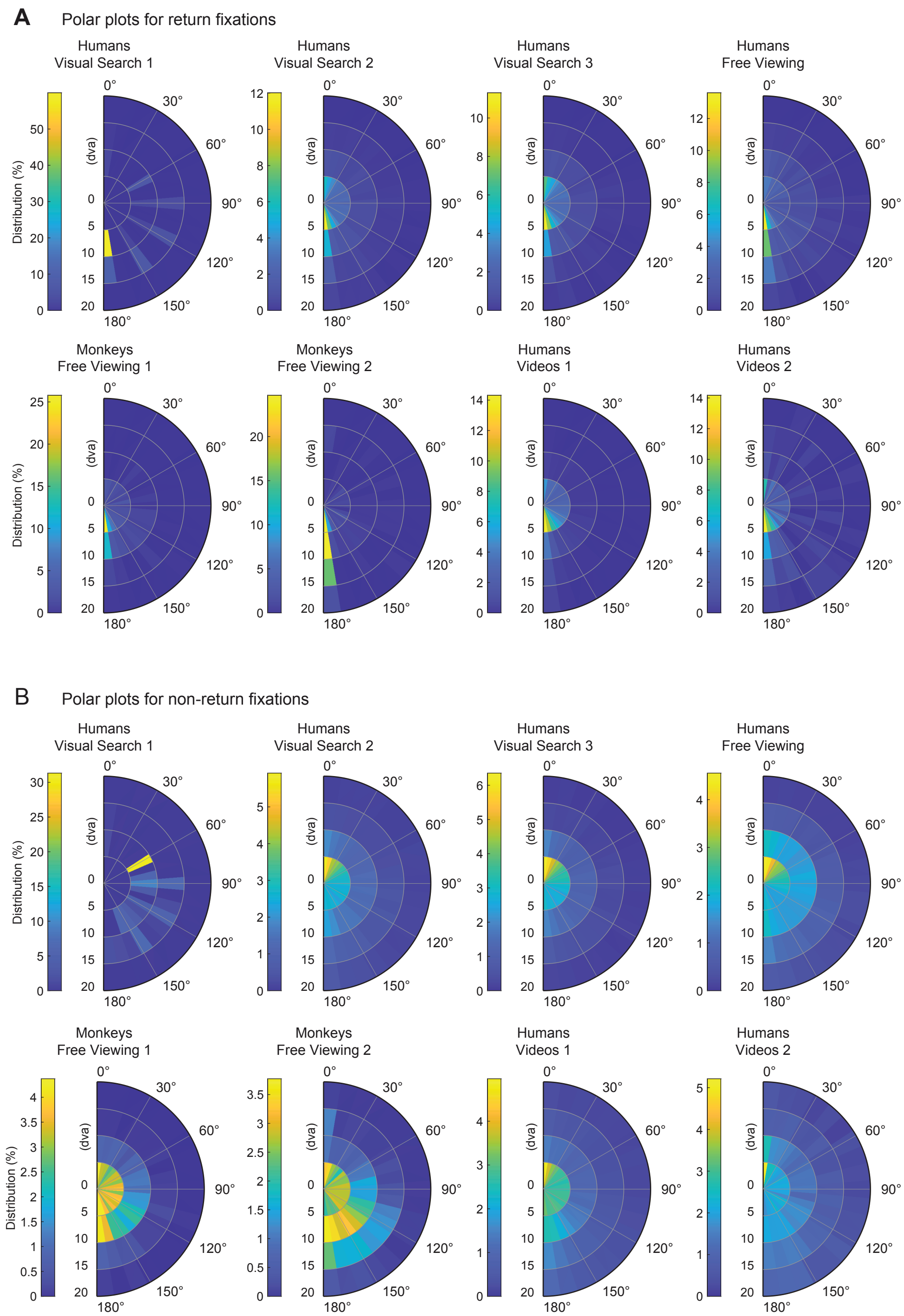


Fig S7. Saccade sizes and angles for primates. Polar plot of distribution of saccade sizes (radial coordinate) and angles for return fixations (**A**) and non-return fixations (**B**) for humans and monkey experiments.

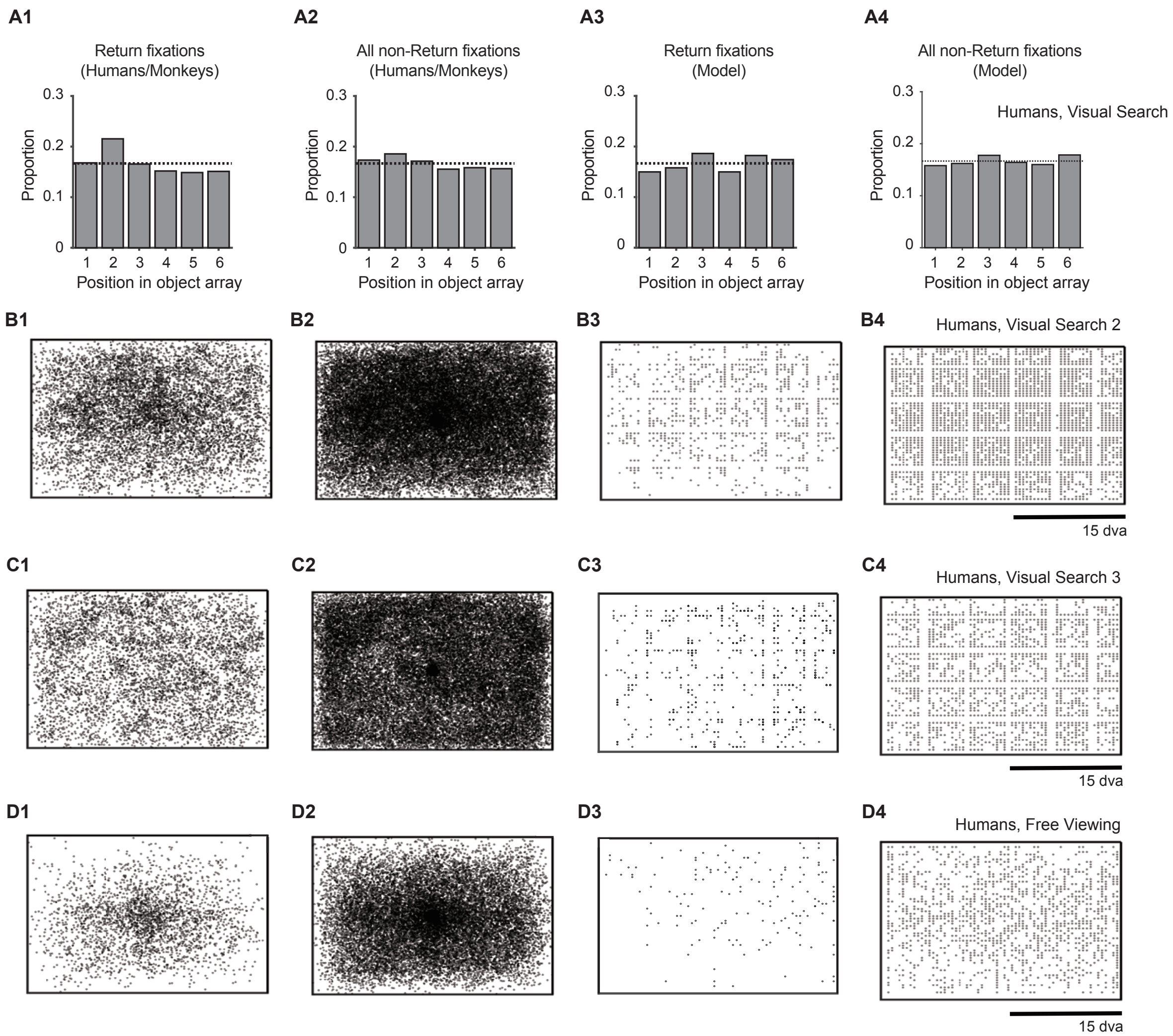


Fig S8. A. Return and non-return fixation locations. Locations of return fixations (Columns 1, 3) and non-return fixations (Columns 2. 4) for each of the experiments for humans/monkeys (Columns 1-2) and the model (Columns 3-4). Each dot indicates the location of a return/non-return fixation.

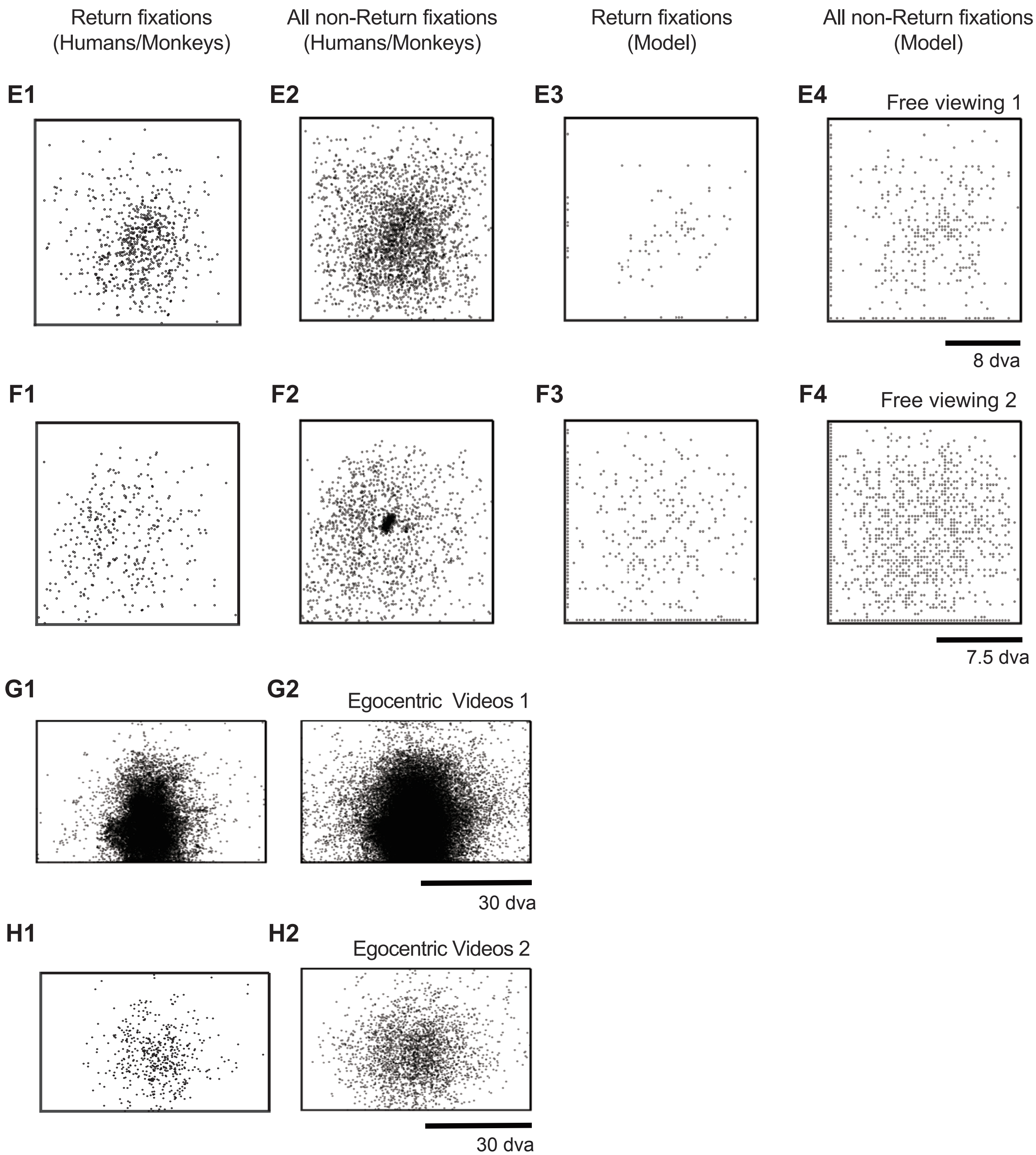


Fig S8. B. Return and non-return fixation locations. Locations of return fixations (Columns 1, 3) and non-return fixations (Columns 2. 4) for each of the experiments for humans/monkeys (Columns 1-2) and the model (Columns 3-4). Each dot indicates the location of a return/non-return fixation.

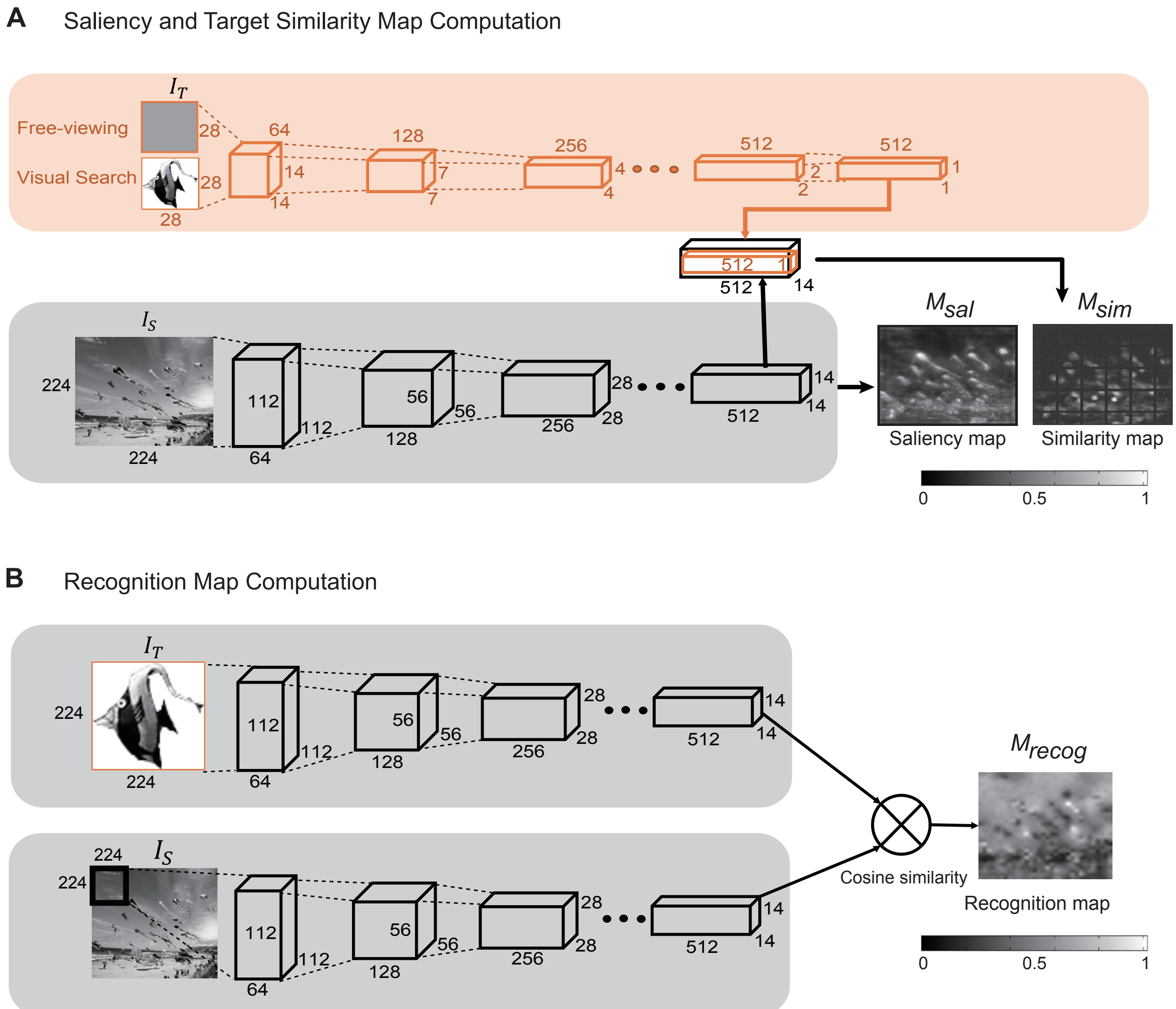


Fig S9. Calculation of bottom-up saliency map, target similarity map, and recognition map in the computational model. (**A**). At the heart of the model is a pre-trained deep convolutional network that mimics image processing along the ventral visual cortex. Here we used the VGG-16 architecture [53]. The features extracted from the top convolutional layer in the model yield the saliency map, M_{sal}. During visual search tasks, we follow the work of Zhang et al [8], to create a target similarity map, M_{sim}, based on comparing the target features (orange box) and the search image features (gray box). Only some of the layers are shown here for simplicity, the dimensions of the feature maps are indicated for each layer. (**B**). For each location on the search image, we cropped a fixation patch of size 224×224 and used it as input to the same pre-trained recognition network as (**A**) to extract the feature maps. Similarly, we obtained the feature maps of the target image. We computed the cosine similarity between the target and cropped features, resulting in a 2D spatial recognition map M_{recog} where each location on the map denotes a confidence value of recognizing the target at each location on I_S.

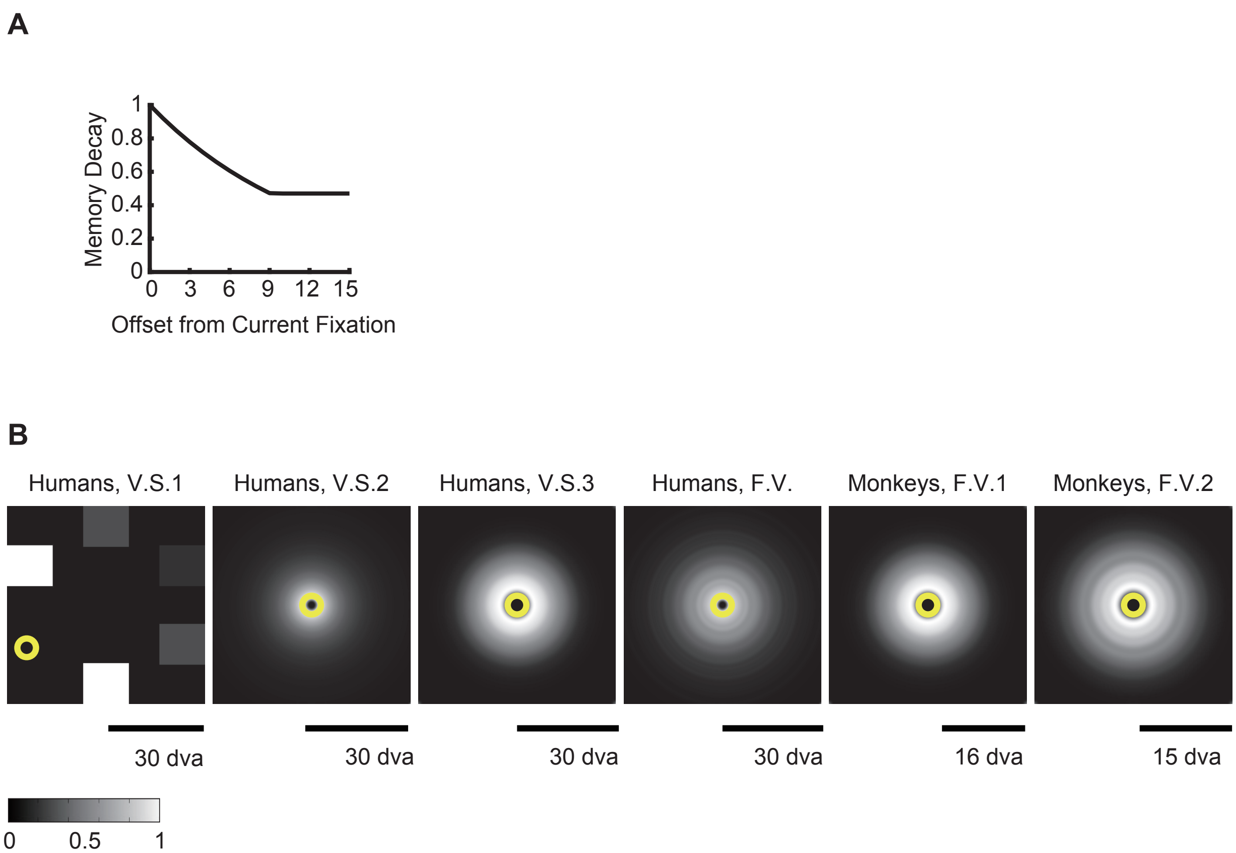


Fig S10. Memory decay function and 2D empirical distribution of saccade sizes. (**A**). The memory decay is a function of the offset from the current fixation (see **Methods**). (**B**). Saccade size prior maps (M_{sac}) for each of the experiemnts. The map activation color scale is shown on the bottom left. In all cases, we show the map assuming fixation in the center (yellow circle), except in the first column where the map is shown assuming fixation in the lower left (yellow circle).

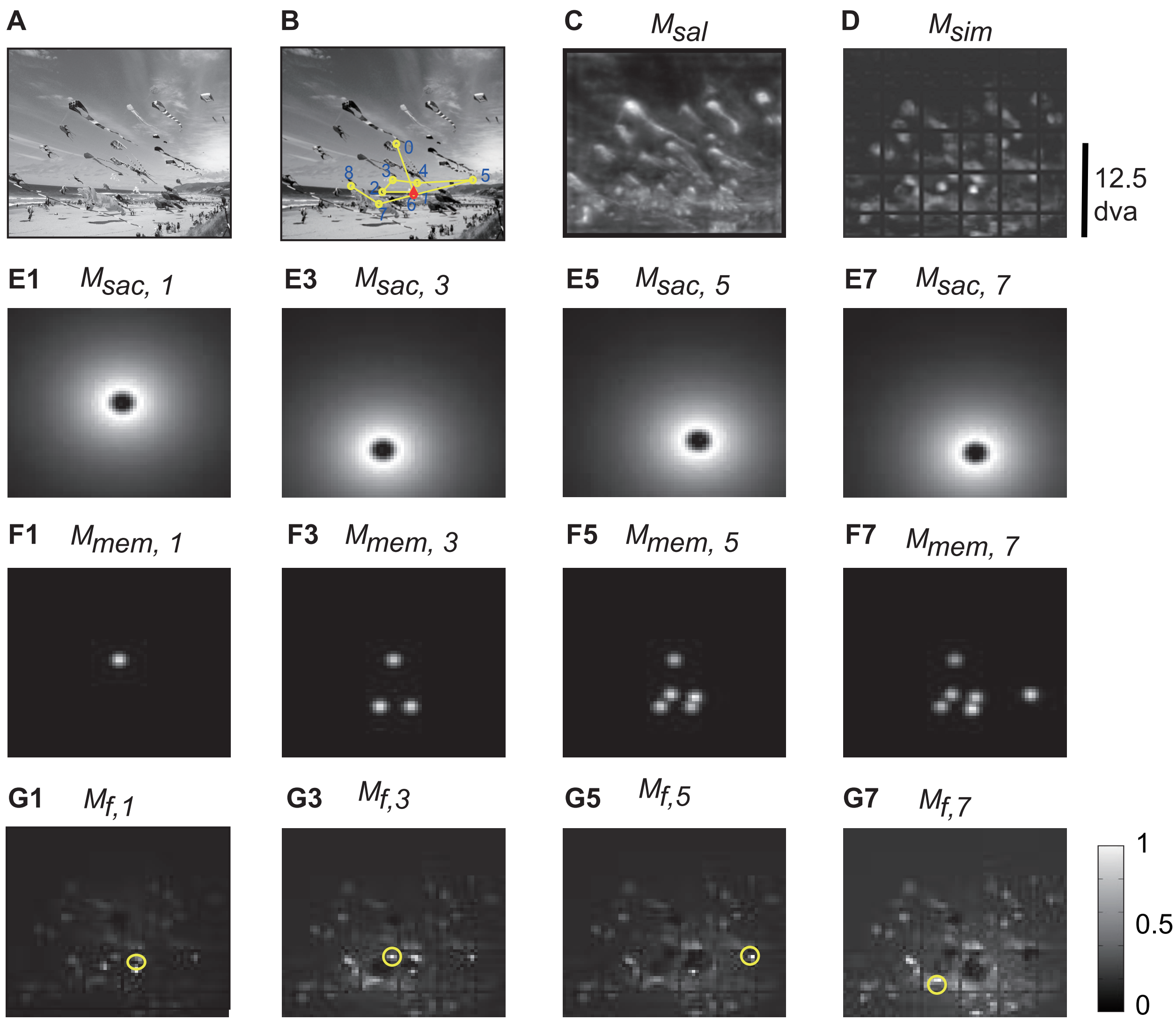


Fig S11. Visualization examples of attention maps for the model. (**A**) Example image. (**B**) Pattern of fixations predicted by the model. The plot follows the conventions in **Fig 1**. (**C**) Saliency map (M_{sal}). (**D**) Similarity map (M_{sim}). ((**E**)) Saccade map (M_{sac}) for fixations 1, 3, 5 and 7. (**F**) Memory map (M_{map}) for fixations 1, 3, 5 and 7. (**G**) Attention map (M_f) for fixations 1, 3, 5 and 7. The yellow circle indicates the maximum, which corresponds to the fixation location.

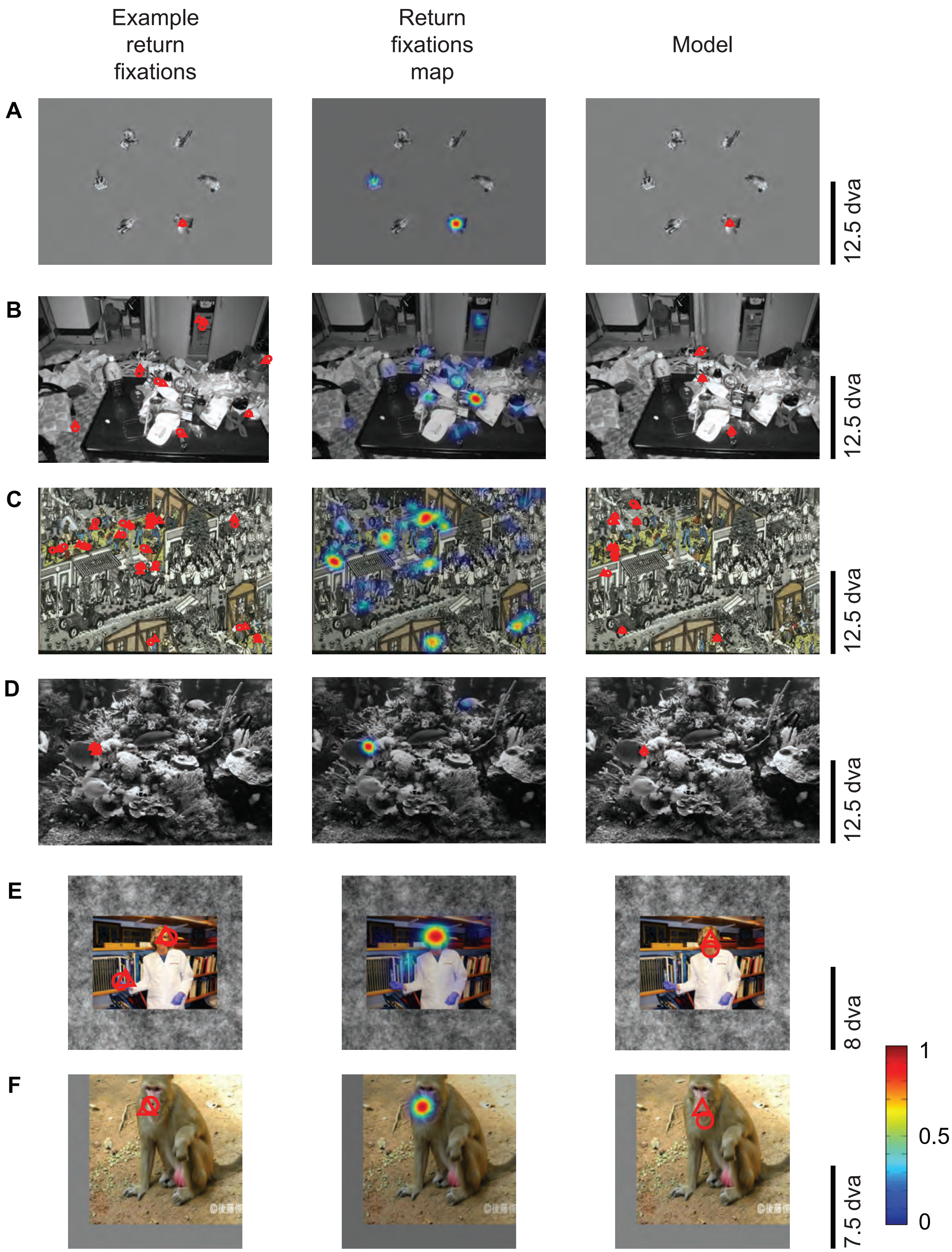


Fig S12. Example return fixations in model predictions. Red-circles = to-be-revisited locations. Red triangles = return fixations. The left column shows data from humans and monkeys, the middle column shows return fixation probability maps across all subjects (see color scale on bottom right), and the rightmost column shows fixations for the model.

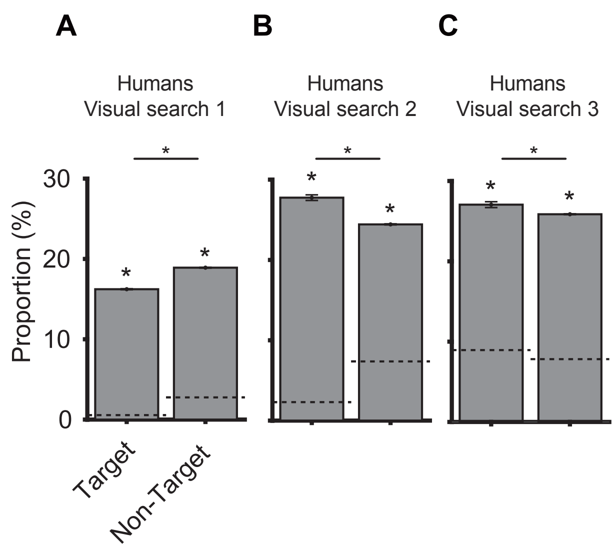


Fig S13. The model makes more return fixations at target locations than non-target locations in visual search. Following the format in **Fig. 6B**, this figure shows the proportion of return fixations in each visual search dataset at target and non-target locations for the model. Asterisks denote statistical significance above chance ($*$ denotes $p < 0.05$, two-tailed t-test).

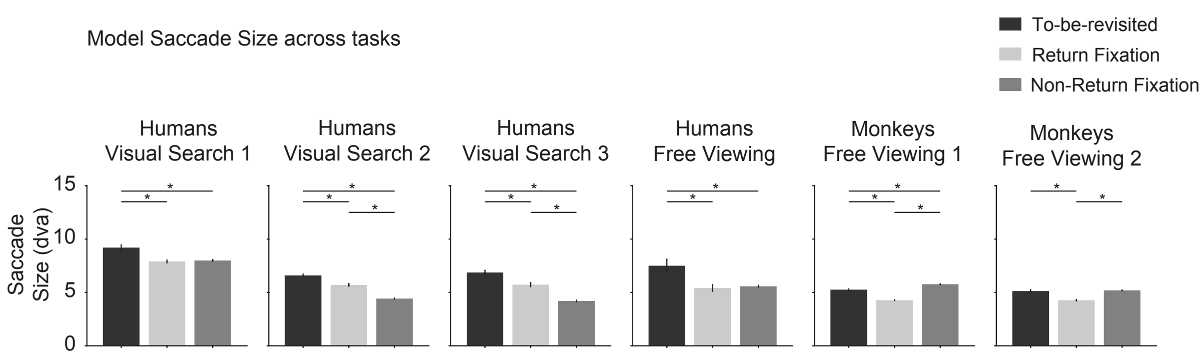


Fig S14. The model makes shorter saccades at return fixations than non-return fixations. Following the format of **Fig. 4B**, bar plots show saccade sizes of to-be-revisited, return, and non-return fixations for the computational model over all static image datasets. Asterisks denote statistical significance ($*$ denotes $p < 0.05$, two-tailed t-test).

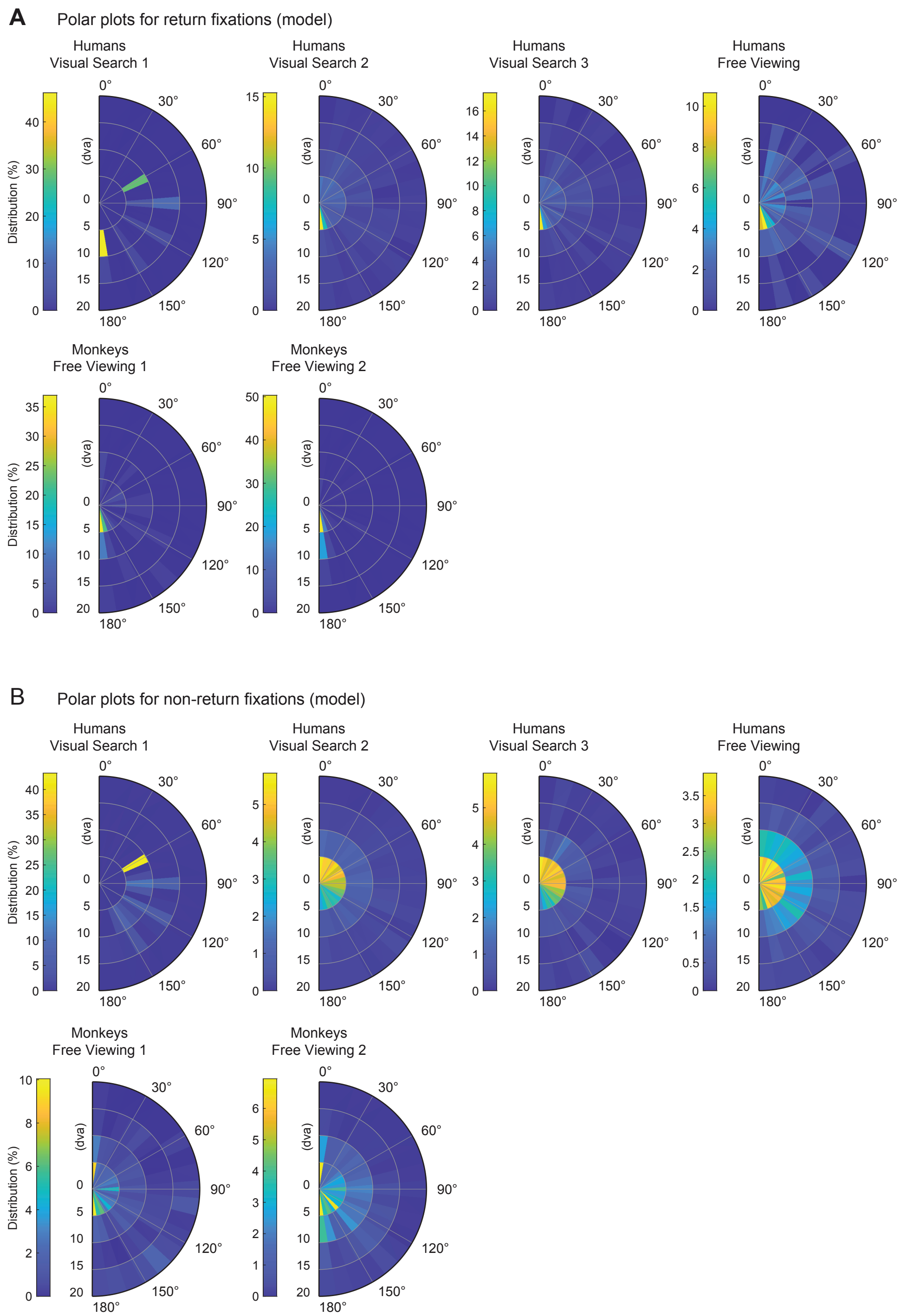


Fig S15. Saccade sizes and turning angles for the model. Polar plot of distribution of saccade sizes (radial coordinate) and turning angles preceding return fixations (**A**) and non-return fixations (**B**) for the model.

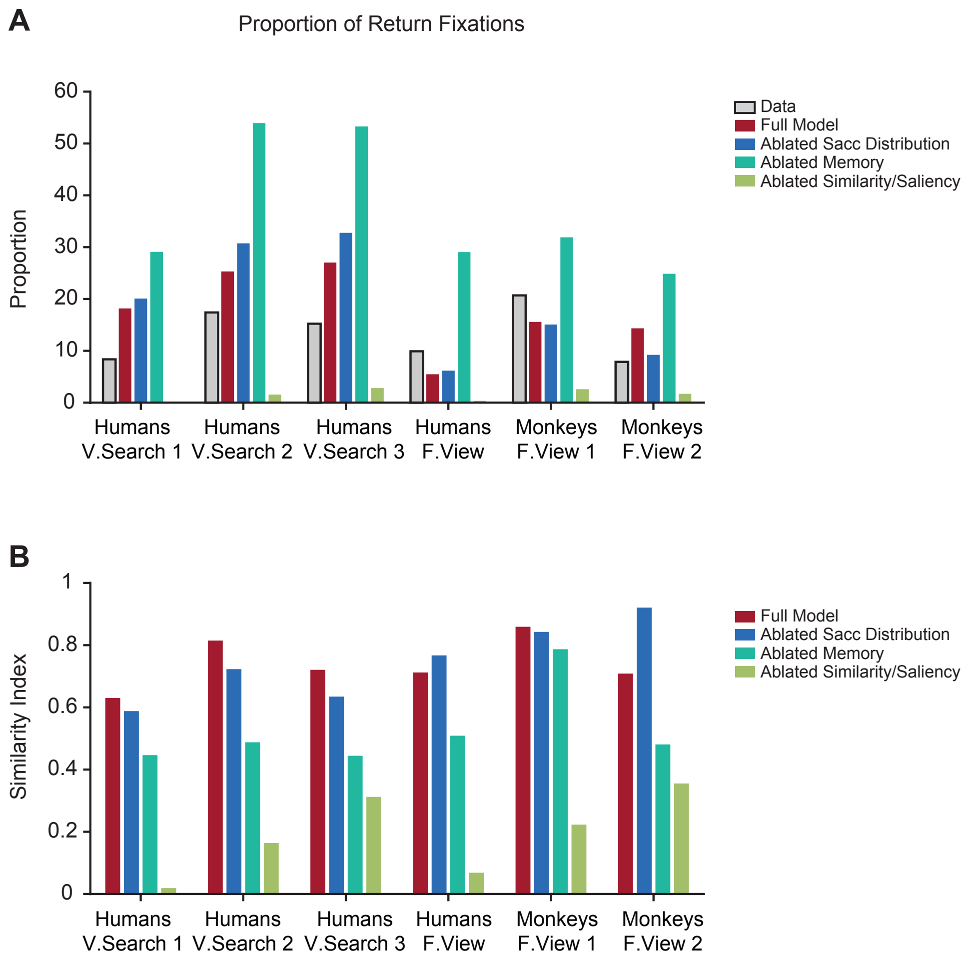


Fig S16. Model ablations reveal critical model components. (**A**). Proportion of return fixations for subjects (gray), full model (red), and three ablated models (see text for details about each ablation). (**B**). Similarity index for return fixations between models and experimental data.

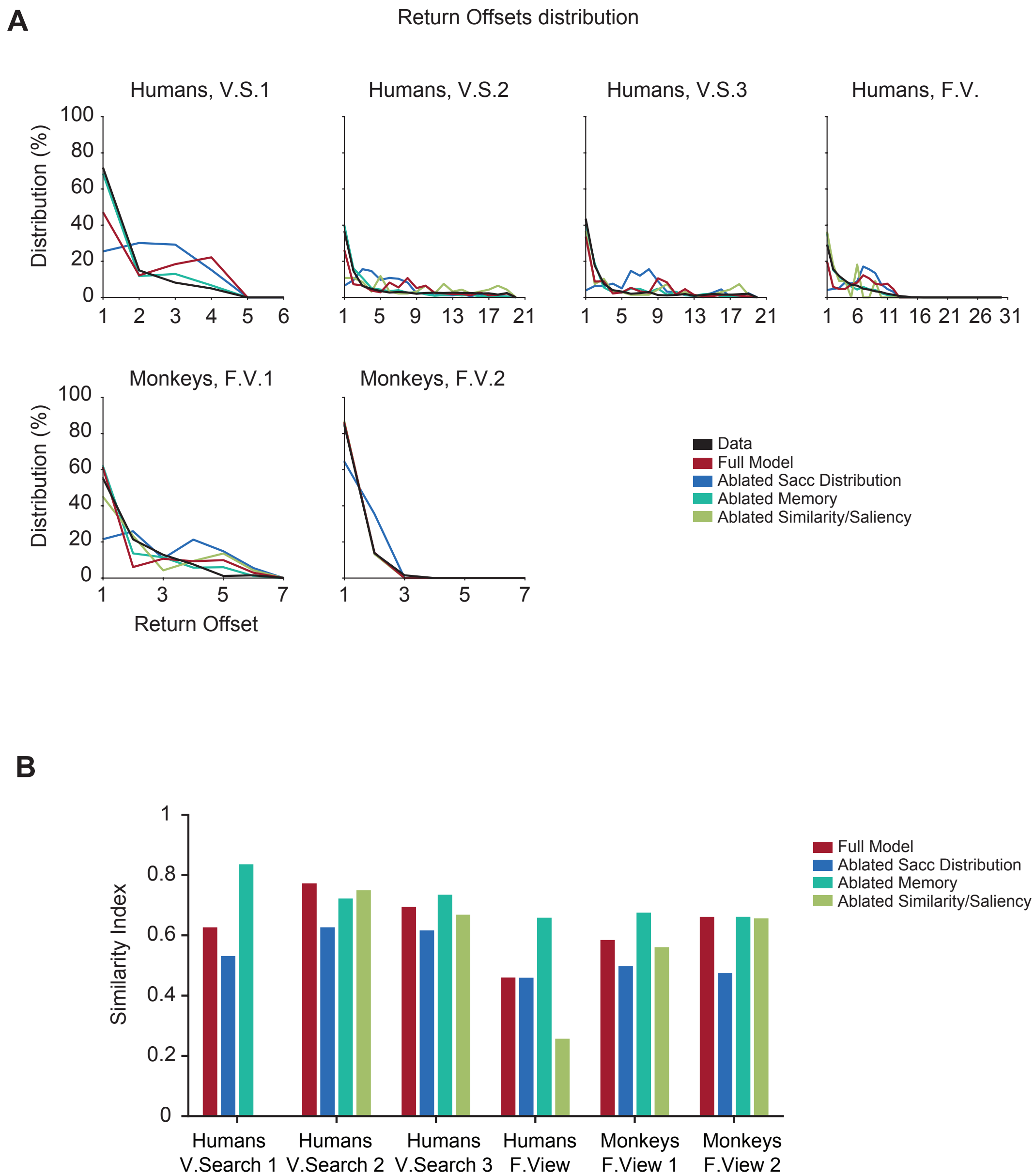


Fig S17. Effect of model ablations on return offset. (**A**). Distribution of return offsets for primates, default model, and ablated models. (**B**). Similarity index for return offsets between models and experimental data.

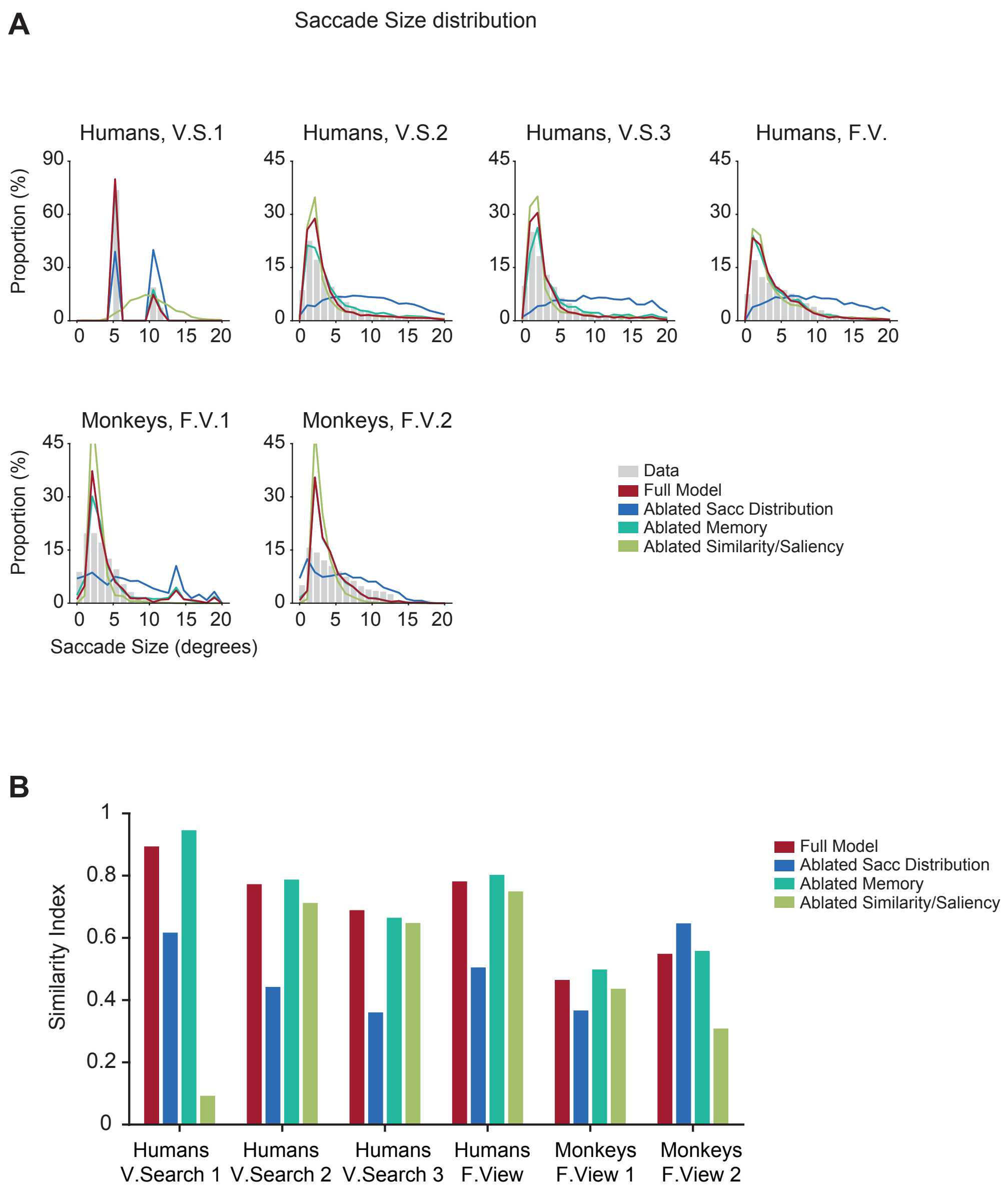


Fig S18. Effect of model ablations on saccade size distribution. (**A**). Saccade size distribution for primates, default model, and ablated models. (**B**). Similarity index for saccade sizes between models and experimental data.

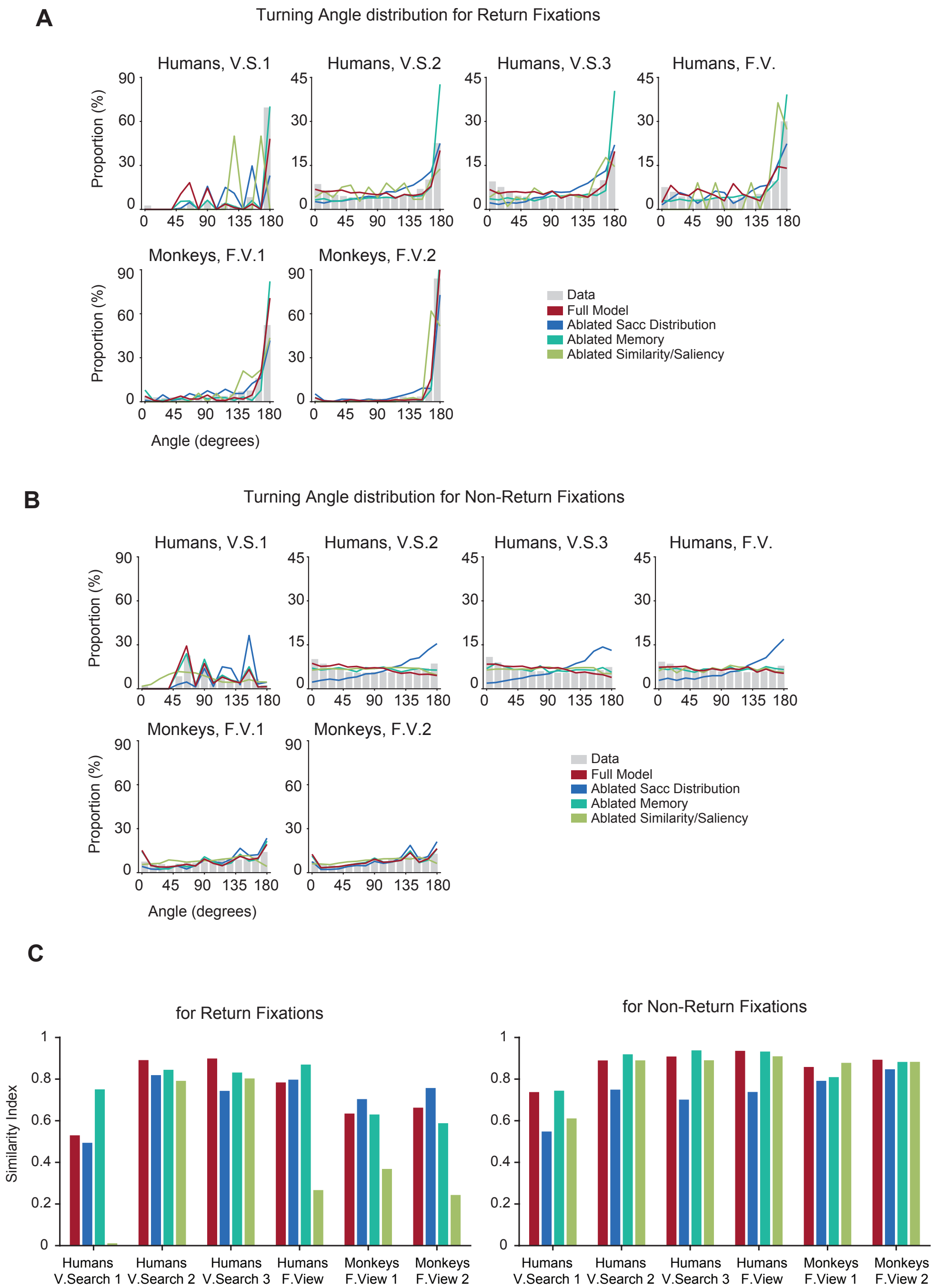


Fig S19. Effect of model ablations on turning angles. (**A-B**). Turning angle distribution for return fixations (**A**) and non-return fixations (**B**) for primates, default model, and ablated models. (**C**). Similarity index for turning angles between models and experimental data preceding return fixations (left) and non-return fixations (right).

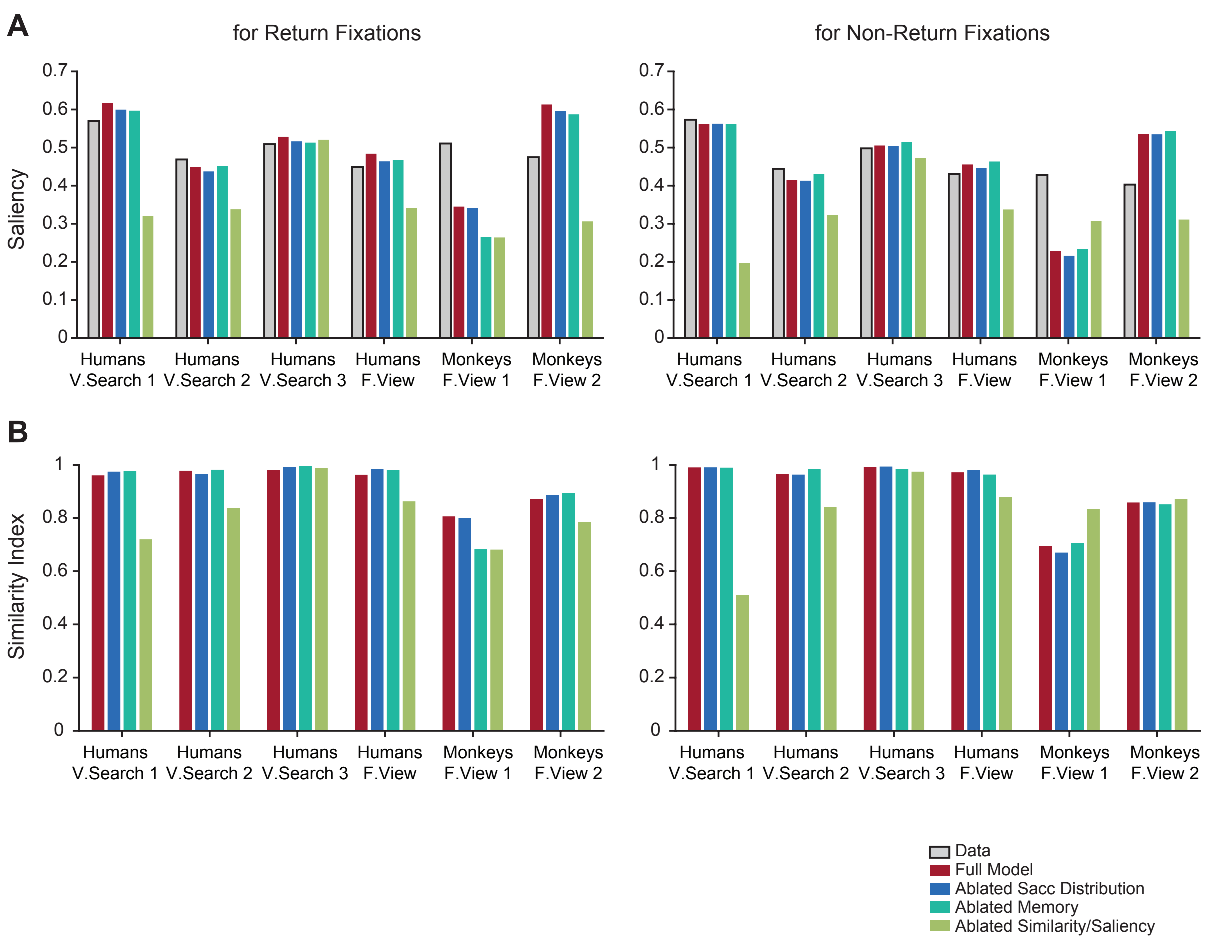


Fig S20. Effect of model ablations on saliency at fixated locations. (**A**). Saliency plot for primates and models at return (left) and non-return (right) fixations. (**B**). Similarity index plot of saliency between primates and all computational models for return (left) and non-return (right) fixations.

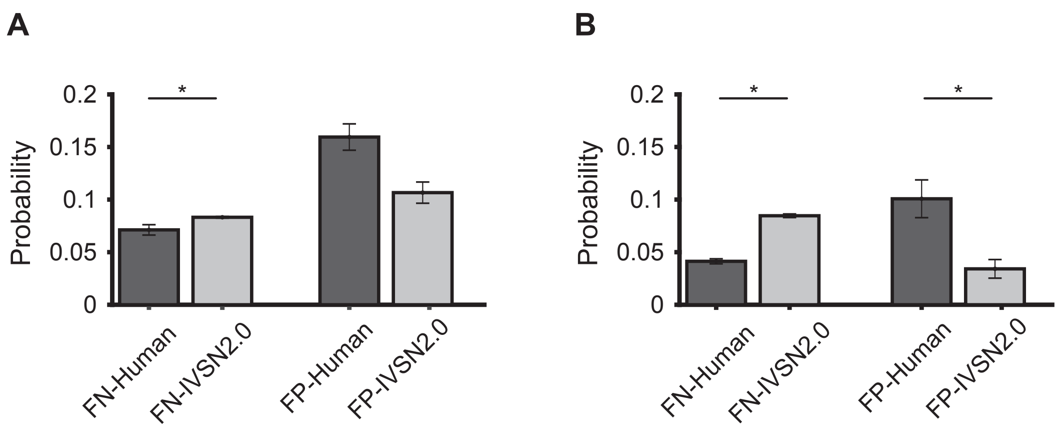


Fig S21. False negative and positive rates for humans and the model in Visual Search 2 on natural images (A) and in Visual Search 3 on Waldo images (B). In visual search tasks, the model has to decide whether the current fixation patch is the target. We introduced the recognition map indicating the confidence value of recognizing that fixation patch belongs to the target (see **Fig. S9B** for recognition map M_{recog} calculation). We empirically set a threshold for deciding whether the target is found based on the corresponding confidence value obtained from M_{recog} at current location l_t. The false positive rates for humans are defined as the number of mouse clicks at the wrong locations normalized by total number of mouse clicks over all the trials. The false negative rates for humans are defined as the proportion of number of fixations falling on the targets and yet, humans do not recognize the targets and continue to move their eyes to other locations out of total number of fixations over all the trials. For the model, since the fixations and the mouse clicks are always consistent, we calculated both false positive rates and false negative rates based on the empirical thresholds and normalized them based on total number of fixations.

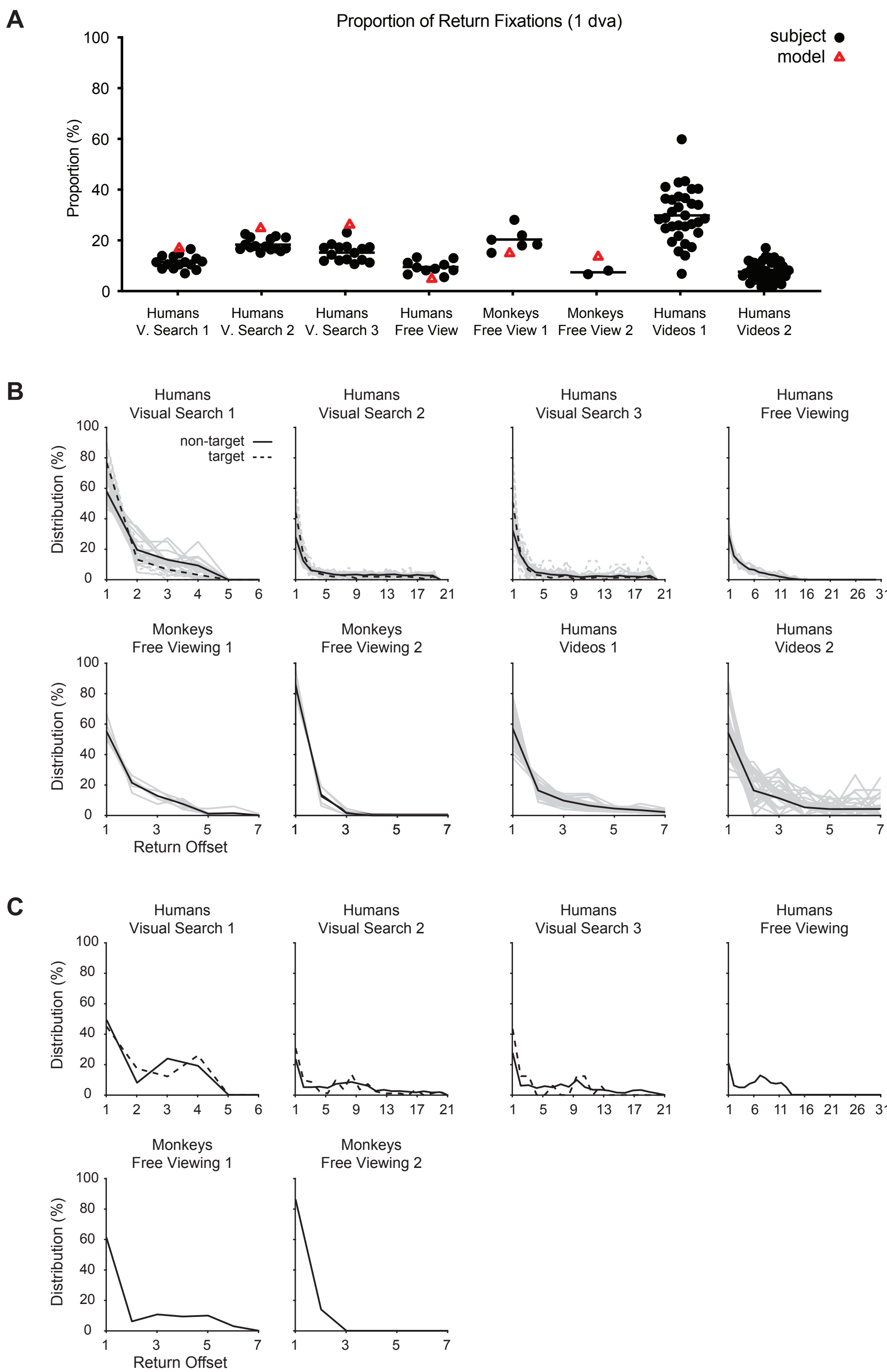


Fig S22. Effect of threshold to define return fixations. We repeated the analyses using thresholds of 1 dva (**A-C**), 2 dva (**D-F**), 3 dva (**G-I**), and 4 dva (**J-L**) to define return fixations. **A, D, G, J**. Proportion of return fixations (format as **Fig. 8B**). **B, E, H, K**. Distribution of return offsets for primates (format as **Fig. 3J**). **C, F, I, L**. Distribution of return offsets for the model (format as **Fig. 8C**).

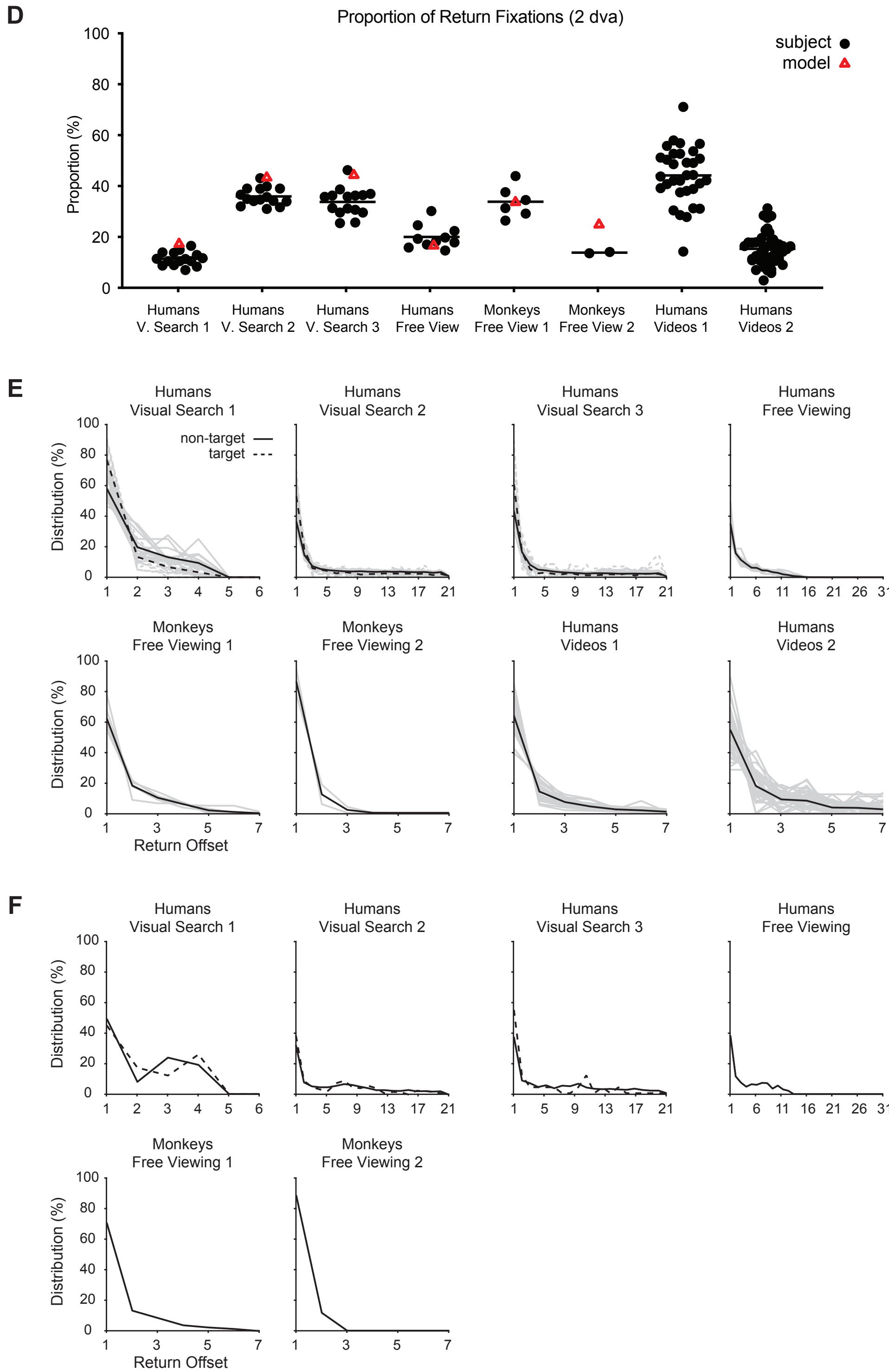


Fig S22. Effect of threshold to define return fixations. We repeated the analyses using thresholds of 1 dva (**A-C**), 2 dva (**D-F**), 3 dva (**G-I**), and 4 dva (**J-L**) to define return fixations. **A, D, G, J**. Proportion of return fixations (format as **Fig. 8B**). **B, E, H, K**. Distribution of return offsets for primates (format as **Fig. 3J**). **C, F, I, L**. Distribution of return offsets for the model (format as **Fig. 8C**).

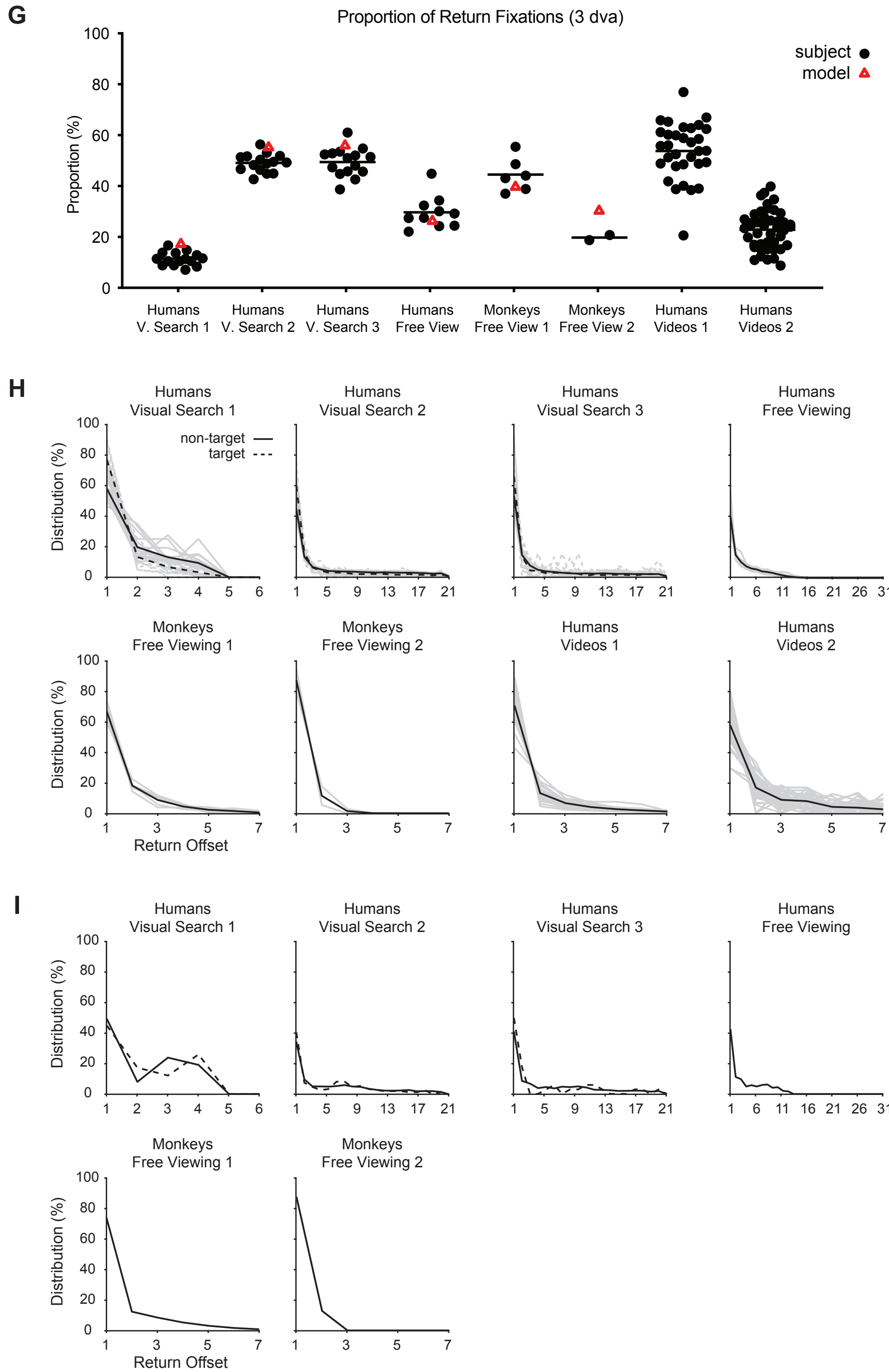


Fig S22. Effect of threshold to define return fixations. We repeated the analyses using thresholds of 1 dva (**A-C**), 2 dva (**D-F**), 3 dva (**G-I**), and 4 dva (**J-L**) to define return fixations. **A, D, G, J**. Proportion of return fixations (format as **Fig. 8B**). **B, E, H, K**. Distribution of return offsets for primates (format as **Fig. 3J**). **C, F, I, L**. Distribution of return offsets for the model (format as **Fig. 8C**).

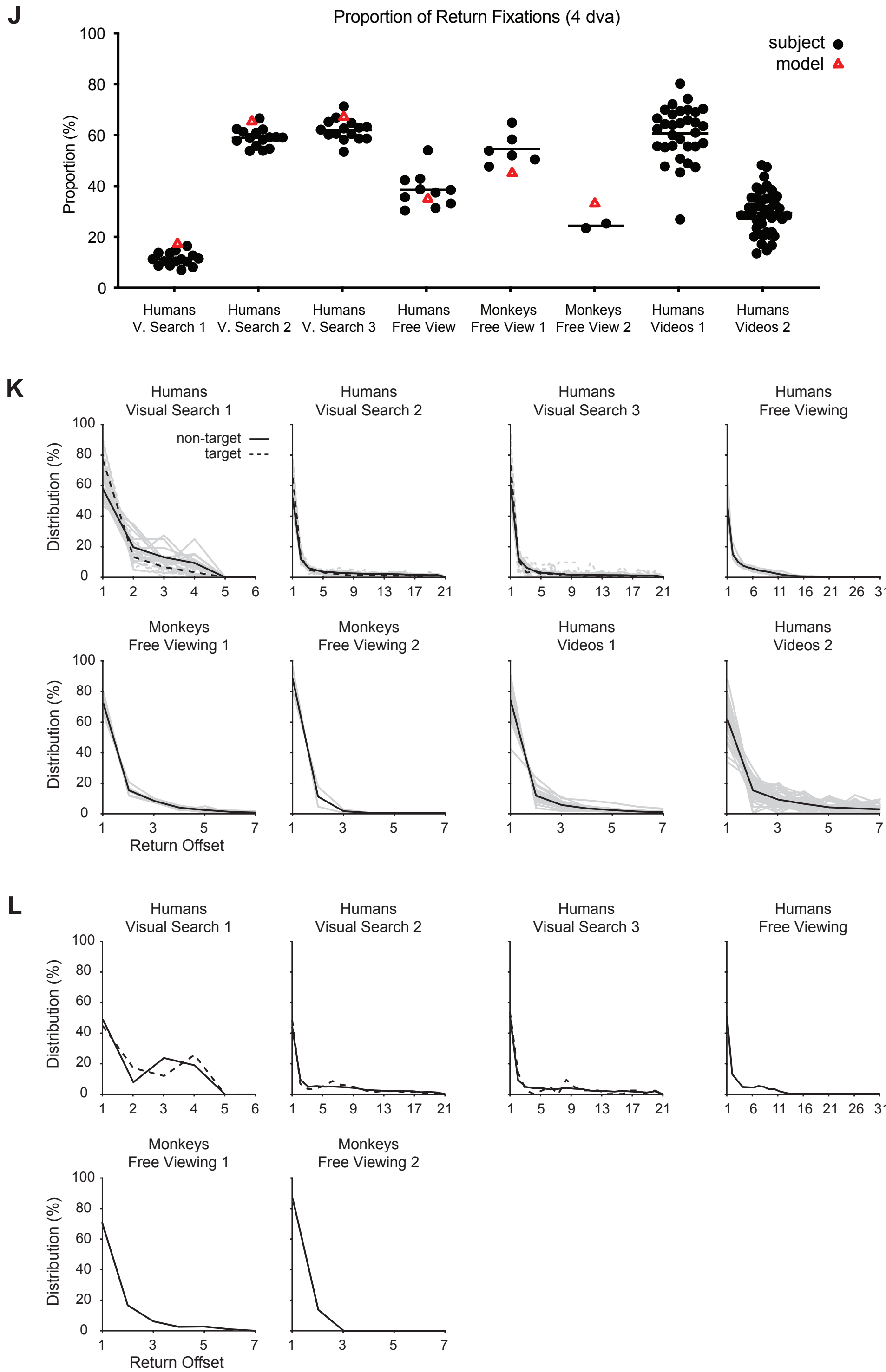
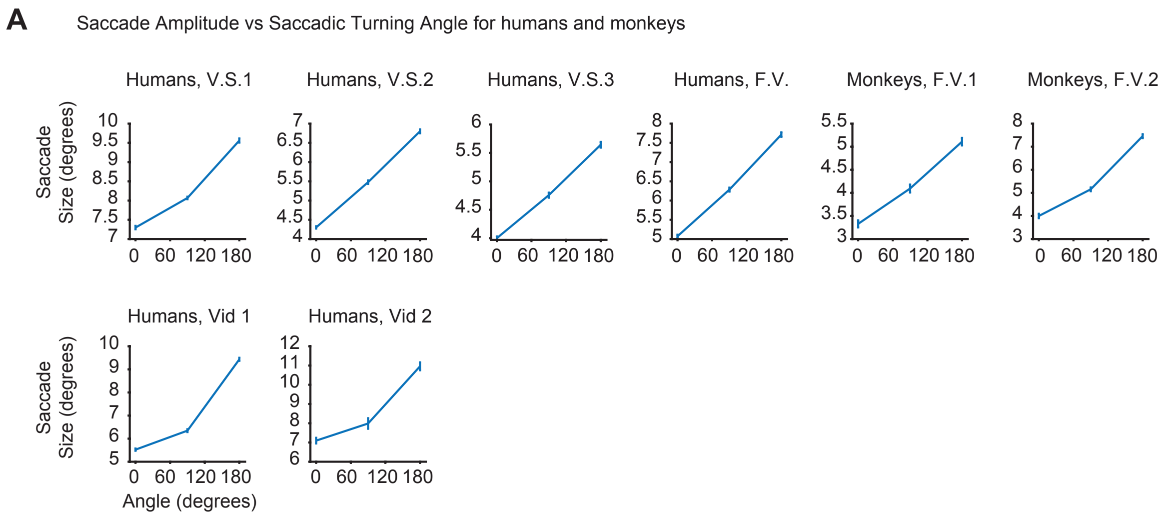


Fig S22. Effect of threshold to define return fixations. We repeated the analyses using thresholds of 1 dva (**A-C**), 2 dva (**D-F**), 3 dva (**G-I**), and 4 dva (**J-L**) to define return fixations. **A, D, G, J**. Proportion of return fixations (format as **Fig. 8B**). **B, E, H, K**. Distribution of return offsets for primates (format as **Fig. 3J**). **C, F, I, L**. Distribution of return offsets for the model (format as **Fig. 8C**).



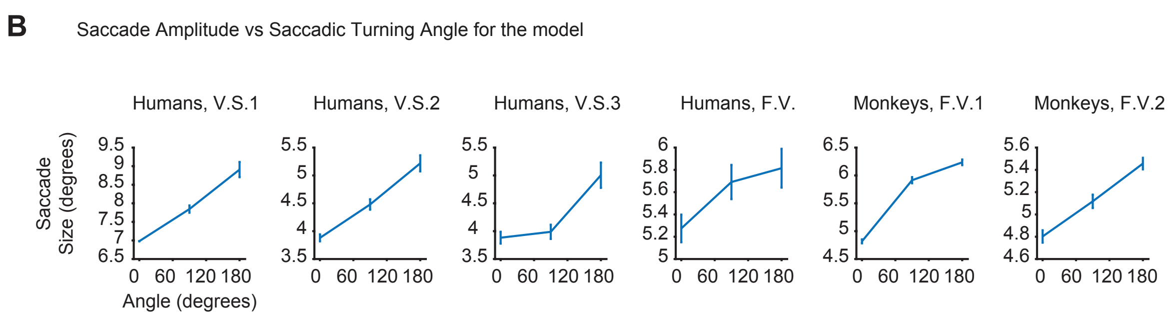


Fig S23. Primates and our model tend to make longer saccades with larger turning angles. Relation between turning angle and saccade size for primates (**A**) and the model (**B**). Both primates and the model showed positive linear correlations between turning angles and saccade sizes.

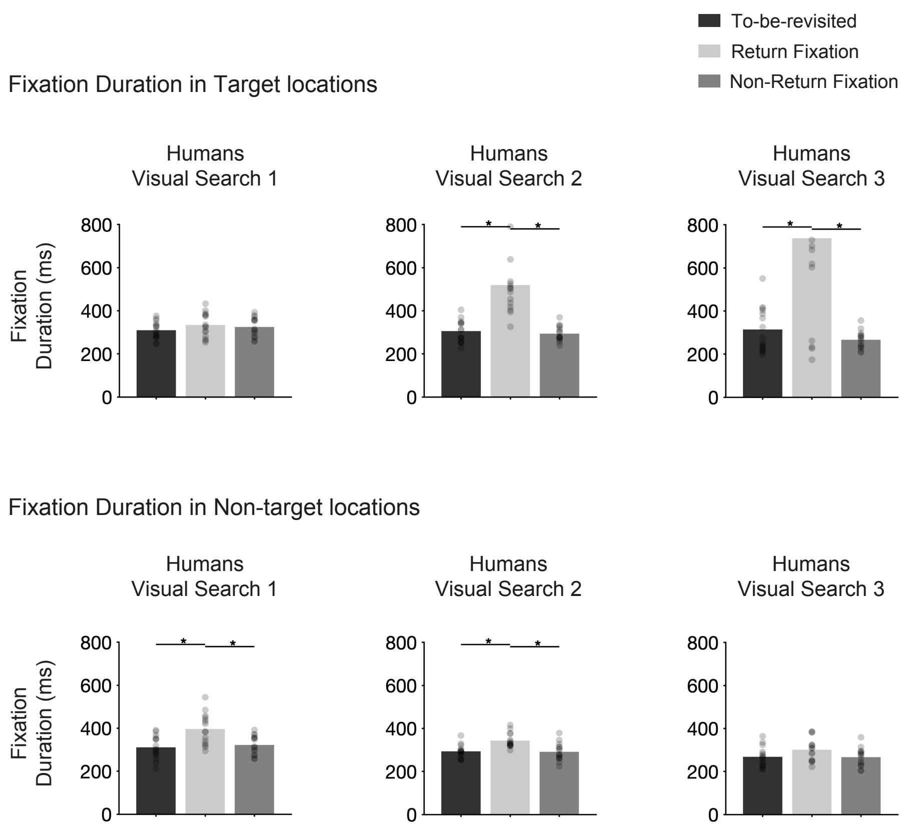


Fig S24. Return fixation durations are longer for target compared to non-target locations. This figure expands upon and follows the same format as **Figure 4A**. Row 1: fixation durations for target locations. Row 2: fixation durations for non-target locations. Only visual search tasks are shown here because there are no target versus non-target locations during the free-viewing experiments.