Neuron, volume *69*

Supplemental Information

Stimulus-Driven Orienting of Visuo-Spatial Attention in Complex Dynamic Environments

Davide Nardo, Valerio Santangelo, and Emiliano Macaluso

Figure S1

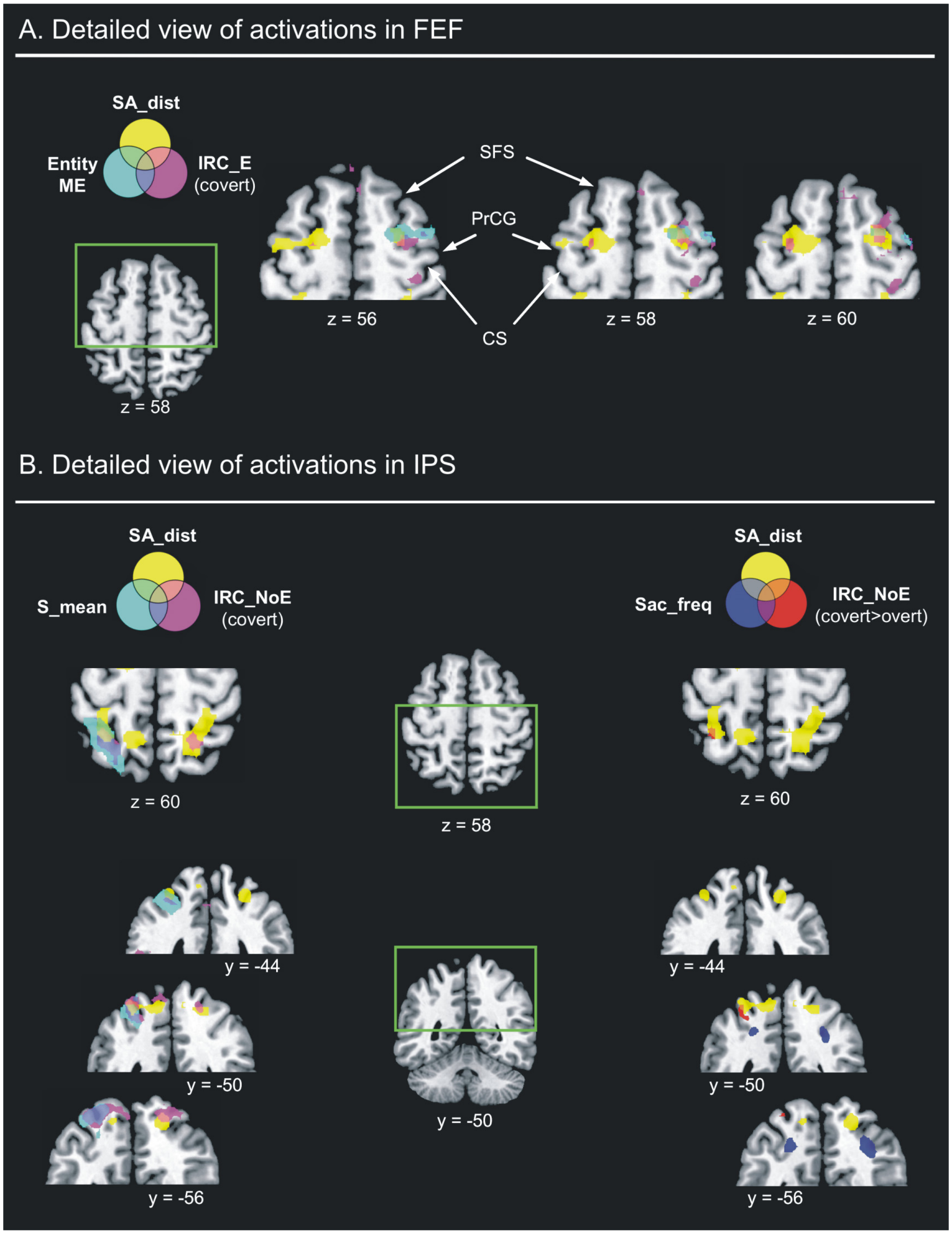

Figure S1: ***Detailed view of activations observed in the dorsal fronto-parietal network, related to Figure 1D.*** **A:** Activations in the frontal-eye fields (FEF) for: distance between maximum salience and attended position (SA_dist); overall response to the human-like characters' appearance (Entity ME); Inter-Runs Co-activation while viewing the *Entity* video in the covert condition (IRC_E). **B:** Activations in the intraparietal sulcus (IPS) and superior parietal gyrus for: distance between maximum salience and attended position (SA_dist); mean saliency (S_mean); Inter-Runs Co-activation while viewing the *No_Entity* video in the covert condition (IRC_NoE); Move forth, after (Sac_freq); Inter-Runs Co-activation while viewing the *No_Entity* video in the covert vs. overt condition (IRC_NoE). SFS: superior frontal sulcus; PrCG: precentral gyrus; CS: central sulcus.

Supplemental Experimental Procedures

Subjects

Eleven volunteers (3 males, range: 25-37, mean age: 29.8) took part in the preliminary behavioural study, involving only eye-movements recordings. Thirteen different right-handed volunteers (6 males, range: 20-39, mean age: 26) underwent the fMRI experiment. All participants were healthy, free of psychotropic or vasoactive medication, with no past history of psychiatric or neurological diseases. All had normal or corrected-to-normal (contact lenses) visual acuity. After having received instructions, all participants gave their written consent. The study was approved by the independent Ethics Committee of the Santa Lucia Foundation (Scientific Institute for Research Hospitalization and Health Care).

Visual stimuli

The *Entity* and *No_Entity* videos were motion-captured by using Pinnacle Studio 11 (*www.pinnaclesys.com/PublicSite/en/Home*) and saved with a resolution of 640 × 448 pixels and a frame rate of 25 Hz. Each video lasted 5 minutes. In the preliminary study, the videos were presented via a computer screen. During fMRI they were back-projected onto a screen at the back of the MR bore that was visible to the subjects via a mirror. In both experiments, the display covered a visual angle of approx. 24 × 16 deg.

Each frame of the *No_Entity* video was analysed using the "SaliencyToolbox 2.2." (http://www.saliencytoolbox.net/; see Walther & Koch, 2006) implemented in MATLAB 7.4. This computes centre-surround contrasts separately for intensity, colour, and orientation in form of conspicuity maps (Itti & Koch, 2001). For each frame, the saliency map was derived from the conspicuity maps by equally weighting each visual feature (all other default parameters were also retained). Each saliency map (40 × 28 pixels) was linearly interpolated to the original image resolution (640 × 448). Saliency maps were used to generate the S_mean (image mean saliency) and the SA_dist ('salience-attention' distance) predictors for subsequent fMRI analyses.

We also evaluated, qualitatively, whether there was some relationship between the points of maximum salience and objects in the scene. In agreement with previous work (Elazary & Itti, 2008; reporting a 43% correspondence between most salient location and a-priori identified objects), we found that the bottom-up saliency algorithm often identified objects as points of maximum salience (e.g. see Fig. 1B, second and third panels). However, it should be noted that models explicitly including object-information can improve fixations prediction compared to bottom-up saliency only, consistent with an additional role of top-down guidance (e.g. Einhäuser et al., 2008; see also Walther & Koch, 2006, who proposed an intermediate approach with the automated extraction of proto-objects from low-level saliency maps).

For the *Entity* video, we extracted manually the horizontal and vertical coordinates of the centre of character's face/head, as this was the location that subjects foveated more systematically (e.g. see Fig. 2B). The average number of frames in which a character was visible was 90 (range 50-155; mean: 3.6 sec., range: 2.0-6.2 sec.). The characters' coordinates were analysed together with the gaze-position data to generate the A_time and A_ampl parameters (i.e. processing time and amplitude of the attentional shifts).

Gaze-position data

In the preliminary behavioural study, horizontal and vertical gaze-positions were recorded with an infrared ASL eye-tracking system operating at 50 Hz (Applied Science Laboratories, Bedford, MA; Model 504). During all fMRI-runs gaze-position was tracked with the same system, now mounting long-range optics for use in the scanner and operating at 60 Hz. This system is fully MR-compatible and does not produce any artifact in the BOLD images. All eye-tracking data were re-sampled at the frame rate of the computer-generated videos (25 Hz).

For the *Entity* video, we compared character-related eye-movements recorded in the preliminary study and during MR scanning ("in-scanner" data) in several ways. First, for each character, we computed the distance between the group-median gaze-position and the character, on a frame-by-frame basis. We correlated these distance-traces in the two groups, revealing significant correlations for 24 out of 25 characters. Next, we set up 25 linear models (one for each character) fitting the "in-scanner" distance-traces with the distance-traces recorded in the preliminary study. The resulting 25 parameter estimates were entered in a one-sample t-test confirming that, on average, the data of the preliminary study successfully predicted the "in-scanner" data. Finally, we applied the three criteria described in the main text (see Experimental Procedures section) to the "in-scanner" gaze-position data. This confirmed as "attention grabbing" 12 out of 15 characters that were identified in the preliminary study. For the 12 characters identified as "attention grabbing" in both groups we correlated the corresponding time and amplitude parameters (A_time and A_ampl, see results section in the main text).

fMRI protocol

Each subject underwent 7 fMRI-runs (see table S1 below). In four runs, subjects were asked to view the videos without moving their eyes (covert orienting conditions). To facilitate compliance with these instructions, a central fixation cross was presented in the centre of the visual display. In the remaining 3 runs they were allowed to move their eyes (free viewing, overt orienting conditions). Our main fMRI analyses concerned the covert viewing conditions, because these ensured that all subjects received the same visual/retinal input. Further, covert viewing of the *Entity* video was always presented during runs 3 and 4, in order to obtain a comparable level of familiarity with the complex environment for all subjects when the characters entered the scene. The *No_Entity* video was presented twice before (runs 1-2: covert and overt viewings, counterbalanced between subjects) and twice after the *Entity* video (runs 5-6: covert/overt viewings, counterbalanced between subjects). The free-viewing *Entity* video was always presented in the last run, in order to minimise the influence of viewing this video on the covert runs that were used for our main fMRI analyses (i.e. runs 3 and 4).

Run	Video	Orienting	fMRI-Analyses
1	NoE	Overt or Covert	SPM covariates; IRC; FC
2	NoE	Overt or Covert	SPM covariates; IRC; FC
3	E	Covert	SPM event-related; SPM modulations; IRC; FC
4	E	Covert	SPM event-related; SPM modulations; IRC; FC
5	NoE	Overt or Covert	SPM covariates; IRC; FC
6	NoE	Overt or Covert	SPM covariates; IRC; FC
7	E	Overt	SPM event-related; SPM modulations; FC

Table S1. Study design and analyses of the 7 runs in the scanner. *Video*: NoE = No_Entity video; E = Entity video; *Orienting*: Overt = viewing with eye-movements allowed; Covert = viewing with central fixation; *fMRI-Analyses*: SPM: Statistical Parametric Mapping (covariates: S_mean, SA_dist, Sac_freq; modulations: A_time, A_ampl); IRC: Inter-Run Co-variation. FC: Functional Coupling with the right Temporo-Parietal Junction.

Magnetic Resonance Imaging

A Siemens Allegra (Siemens Medical Systems, Erlangen, Germany) 3T scanner equipped for echo-planar imaging (EPI) was used to acquire functional magnetic resonance (MR) images. A quadrature volume head coil was used for radio frequency transmission and reception. Head movement was minimised by mild restraint and cushioning. Thirty-two slices of functional MR images were acquired using blood oxygenation level-dependent imaging (3 x 3 mm, 2.5 mm thick, 50% distance factor, repetition time = 2.08 s, time echo = 30 ms), covering the entirety of the cortex.

fMRI data pre-processing

Data pre-processing was performed with SPM8 (Wellcome Department of Cognitive Neurology) as implemented on MATLAB 7.4. A total of 592 fMRI volumes for each subject were analysed (4 sessions × 148 volumes). After having discarded the first 4 volumes of each session, images were realigned in order to correct for head movements. Slice-acquisition delays were corrected using the middle slice as a reference. Images were then normalised to the MNI EPI template, re-sampled to 2 mm isotropic voxel size and spatially smoothed using an isotropic Gaussian kernel of 8 mm FWHM (full with half maximum).

Additional fMRI analyses using "in-scanner" indexes of orienting efficacy

The recording of eye-movements data during fMRI (overt viewing fMRI-runs, see also Table S1, above) enabled us to analyse covert viewing fMRI data using behavioural indexes derived from the same group of subjects ("in-scanner" indexes). Moreover, because our experimental approach comprised unrepeated and complex stimuli that will inevitably result in poorer signal-to-noise ratio compared with standard fMRI paradigms, we also tested all attention-related effects using more targeted subject-specific ROIs in the dorsal and ventral attentional systems.

All first-level within-subject models (*No_Entity* and *Entity* videos) were re-constructed, now using predictors based on the "in-scanner" parameters. Group-level statistics consisted of one-sample t-tests and a full-factorial ANOVA (cf. also main text) re-assessing all attention-related effects (S_mean, SA_dist and Sac_freq, for the *No_Entity* video; main effect of characters' onset, attention "grabbing vs. non-grabbing" characters, A_time and A_ampl for the *Entity* video). We examined activity in four ROIs belonging to the dorsal fronto-parietal network (aIPS and FEF, bilaterally) and four ROIs in the ventral fronto-parietal network (TPJ and IFG, bilaterally). MarsBar

0.41 (MARSeille Boîte À Région d'Intérêt, SPM toolbox) was used to extract and average data across all voxels in each ROI.

In the dorsal fronto-parietal network, the four ROIs were defined individually in each subject (spheres with a radius of 8 mm). ROIs were centred considering the subject's peak activation within the clusters showing an effect of SA_dist in the whole-brain group analysis (see Fig. 1D). Further, we anatomically constrained the centre of each ROI by making sure that the aIPS-ROIs were in the postcentral gyrus or the superior parietal lobule; and that the FEF-ROIs were in the precentral gyrus or the superior frontal gyrus (AAL atlas; Tzourio-Mazoyer et al., 2002).

The ROI analyses of the *No_Entity* video confirmed our main finding about the efficacy of salience for covert spatial orienting in all four dorsal ROIs (SA_dist: all $p < 0.001$). As in the main analyses, none of these areas showed an overall effect of attention shifting (Sac_freq). The test for the overall effect of salience (S_mean) confirmed our finding in the left aIPS ($p < 0.010$) and revealed some modulation also in the FEF ($p < 0.029$ and $p < 0.021$, in the left and right hemisphere respectively). During viewing of the *Entity* video, these more focused ROI analyses revealed an overall response to the characters' appearance in the right aIPS ($p < 0.035$) and in the right FEF ($p < 0.004$, also found in the original whole brain analyses, see Fig. 3A and Fig. S1 panel A), but these were not modulated by the attention-grabbing efficacy of the characters.

In the ventral fronto-parietal network, we defined subject-specific ROIs in the right TPJ (spheres with a radius of 8 mm). These were centred at subject's peak activation within the cluster showing a main effect of characters' onset at the group-level (see Fig. 3A). Again, anatomical constraints ensured that the rTPJ-ROIs were in superior temporal gyrus or supramarginal gyrus (AAL atlas; Tzourio-Mazoyer et al., 2002). The left TPJ and bilateral IFG did not show any significant activation in our main analyses (cf. Figs. 1 and 3). Therefore, the left TPJ-ROI was created selecting voxels in the left hemisphere at symmetric coordinates with respect to the rTPJ-ROI. For the IFG-ROIs, we identified a cluster in the right hemisphere showing significant functional connectivity with rTPJ (see also Fig. 4C). The corresponding left IFG-ROI was created selecting voxels at symmetric coordinates in the left hemisphere.

In the rTPJ-ROI, analyses of the *Entity* video confirmed the overall effect of characters' appearance ($p < 0.001$) and the modulation according to attention-grabbing efficacy of the characters (AG vs. NoAG: $p < 0.037$). The additional temporal and spatial parameters did not reach statistical significance (A_time: $p = 0.151$; A_ampl: $p = 0.318$), but note that these are very subtle parameters now computed on the third presentation of the *Entity* video (fMRI-run 7, see Suppl. Tab. above). In the left TPJ we found some character-related activation ($p < 0.011$), but this was not modulated by attention-grabbing efficacy. The left IFG did not show any response to characters' onset. By contrast, in the right IFG we found significant effects for all tests related to the characters' presentation and their attention grabbing-efficacy (characters' onset: $p < 0.004$; AG vs. NoAG: $p < 0.010$; A_time: $p< 0.004$; A_ampl: $p < 0.001$). Analyses of the *No_Entity* video, confirmed the absence of any effect of mean saliency (S_mean), efficacy of saliency (SA_dist), or overall attention shifting (Sac_Freq) in all ROIs of the ventral fronto-parietal system.

In summary, these additional analyses using subject-specific ROIs and indexes of attention-efficacy derived from "in-scanner" eye-movements data confirmed our main findings showing that the efficacy of salience modulates activity in the dorsal fronto-parietal network (SA_dist, *No_Entity* video); while the efficacy of the human-like characters modulates activity in right TPJ (attention "grabbing vs. non-grabbing" characters, *Entity* video). These more targeted analyses also revealed an overall effect of mean salience in the FEF (*No_Entity* video); and characters-related efficacy in the right IFG (*Entity* video).

<u>Inter-Run Co-variation Analysis (IRC)</u>

The IRC approach is conceptually related to the Inter-Subject Correlation (ISC) first proposed by Hasson and colleagues (Hasson et al., 2004) and is based on the idea that when exposed twice to the same dynamic input, the brain areas that processed the input should show similar patterns of activity over time ("synchronisation"). This has been successfully applied to fMRI data collected during natural vision revealing consistent brain activity across subjects (i.e. high ISC) in areas processing high-order visual information (Hasson et al., 2004), as well as in areas associated with the encoding of episodic memory (Hasson et al., 2008). However, IRC differs from ISC in several respects. First, by computing "synchronisation" within rather than between subjects, IRC minimises the contribution of any functional/anatomical between-subjects variability. Second, the use of co-variation rather than correlation coefficients means that IRC assesses the amount of the "synchronisation", rather than its significance. This has important implications for any statistical test comparing "synchronisation" between conditions (e.g., here within-subjects ANOVA) that, in the case of IRC, will assess any change of synchronisation-level rather than some unspecified combination of changes of synchronisation-level and changes of noise-level. Third, the use of the General Linear Model to estimate within-subject IRC parameters enables removing variance components of no interest. Our regression models included head-movements, global signal and, for the covert viewing conditions, losses of fixation as covariates of no interest. Moreover, the IRC regression model for *Entity* video also included the predicted BOLD response for the human-like characters (i.e. transient activation time-locked to characters onset). This effectively ensured that any significant IRC found specifically for the *Entity* video does not simply reflect a common/consistent activation in response to the appearance of these stimuli (see also Hasson et al., 2008, about the influence of stimulus-related activation on ISC).

Group-level statistical inference concerning the IRC was performed using standard parametric analyses (ANOVA) that require data normality. Accordingly, we verified that the subject-specific parameter estimates of the IRC complied to normality using the Lilliefors test as implemented in Matlab 7.4. This test was chosen because of the relatively small number of observations (n = 13). Tests were performed at each and every voxel separately for the three conditions submitted to the IRC analysis (covert viewing the *Entity* and *No_Entity* videos, plus overt viewing of the *No_Entity* video). The null-hypothesis of data normality was rejected for:

5.79% of the voxels in the covert *Entity* analysis, 6.36% in the covert *No_Entity* analysis, and 5.80% in the overt *No_Entity* analysis. This is consistent with the chosen alpha level = 0.05, indicating that the distributions of the IRC parameter estimates conform to normality.

Supplemental References

Hasson, U., Furman, O., Clark, D., Dudai, Y., and Davachi, L. (2008). Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. Neuron 57, 452–462.

Itti, L., and Koch, C. (2001). Computational modelling of visual attention. Nat. Rev. Neurosci. 2, 194–203.

Walther, D., and Koch, C. (2006). Modeling attention to salient proto-objects. Neural Netw. 19, 1395–1407.