Supplemental Material

Supplementary Methods

Balancing the Complexity of the Immersive Videos Between Experimental Conditions. We adopted the same procedure to ensure that any observed behavioural or fMRI effects of experimental condition would not be due to differences in the complexity of the individual videos as described in our previous study (Iriye & Ehrsson, 2022). Two independent raters judged each video in terms of multiple sub-categories of complexity and composite scores of each sub-category were used to divide videos into two groups (see Supplementary Table 1) to minimize differences in average complexity between groups (see Supplementary Table 2; ; Bonasia, Sekeres, Gilboa, Grady, Winocur, and Moscovitch, 2018; Iriye and Ehrsson, 2022; Sekeres, Bonasia, St-Laurent, Pishdadian, Winocur, Grady, and Moscovitch, 2016). Assignment of each video to an experimental condition was counterbalanced across participants. The individual raters judged each video in terms of its visual, audio, narrative, and emotional complexity, using five-point Likert scales (One = Low, Five = High). Visual complexity was measured along three dimensions concerning complexity of the background scene, amount of movement, and number of characters. Auditory complexity referred to the complexity of the background audio, while narrative complexity referred to complexity of the storyline. Emotional complexity was characterized according to degree of sadness, excitement, joy, anger, disgust, fear, and shame present within each video. A two-way random effects model was run for each complexity dimension to assess inter-rater reliability. The emotional sub-categories of sadness, anger, disgust, fear, and shame did not display enough variance to run the analysis as all ratings from both raters were at floor level (i.e., one). The average intraclass correlation coefficient across complexity categories was 0.78 (*SD* = .08), demonstrating a reliable degree of agreement between raters (see Supplementary Table 3 for means from each sub-category).

Supplementary Table 1.
Average Complexity Scores Per Sub-Category According to Group Assignment

Video Title	Visual	Auditory	Narrative	Emotional	Group
Moderna Museet	2,00	2,50	1,50	1,36	1
Fotografiska	1,83	1,00	2,00	1,93	1
Storkyrkan	2,83	3,00	2,00	1,00	1
Solna Centrum	2,50	1,50	2,00	1,29	1
Vasa Museet	2,67	2,50	2,00	1,21	1
Gamla Stan	3,50	3,00	2,50	1,64	1
Kungtradgården	2,00	2,00	2,50	1,00	1
Stadshuset	2,33	3,00	2,50	1,14	1
Hagaparken	2,67	1,00	2,50	1,14	1
Odenplan	2,83	4,00	3,00	1,07	1
Karlaplan	2,33	1,50	3,00	1,07	1
FilmstadenSergel	2,67	3,00	3,50	1,21	1
Central Station	3,00	2,00	1,00	1,93	2
KI Sjukhuset	4,50	2,50	1,50	2,43	2
Humlegården	1,67	1,50	2,00	1,07	2
Royal Palace	2,17	3,00	2,00	1,00	2
Strömkajen	2,50	3,00	2,00	1,29	2
Skansen	2,33	1,50	2,50	1,14	2
Stadsbiblioteket	1,50	1,00	2,50	1,43	2
Skinnarviksberget	1,83	3,50	2,50	1,21	2
Nytorget	2,33	3,00	2,50	1,07	2
Vasaparken	2,33	2,50	2,50	1,00	2
Globen	3,00	3,00	3,50	1,14	2
Grona Lund	2,83	3,00	3,50	1,14	2

Supplementary Table 2.
Differences in Average Complexity Between Groups of Video Stimuli

	Visual	Auditory	Narrative	Emotional
Group 1	2,51	2,33	2,42	1,26
Group 2	2,00	2,46	2,33	1,32
Difference	0,01	-0,13	0,08	-0,07

Supplementary Table 3.

Average Intraclass Correlation Coeffiecients of Complexity Sub-Categories

Visual Complexity Measure	**Intraclass Correlation Coefficient**
Background	0,80
Movement	0,73
Number of Characters	0,82
Auditory Complexity Measure	
Background Audio	0,88
Narrative Complexity Measure	
Storyline	0,62
Emotional Complexity	
Excitement	0,81
Joy	0,77
Sadness	not enough variance
Anger	not enough variance
Disgust	not enough variance
Fear	not enough variance
Shame	not enough variance

Assessing Presence Within the Immersive Videos. The presence questionnaire was included to monitor the overall feeling that participants experienced "being there" inside the immersive 3D scene, which we reasoned might have unintended effects on participants' ability to recall specific details about each scene, independent from the bodily illusion induction. The presence questionnaire was comprised of three statements that participants rated on seven -point Likert Scales adapted from Slater (2000; see Supplemental Table 4).

Supplementary Table 4. ***Presence Questionnaire***	
To what extent did you have the sense of being in the video?	1 2 3 4 5 6 7 Not at all / Minimally / Moderately / Very Much So
To what extent were there times during the experience when the video became 'reality' for you, and you almost forgot about the real world of the lab in which the whole experience was taking part?	1 2 3 4 5 6 7 None / Some / Several / Many
When you think back about your experience, do you think of the video more as images you saw or more as a location you visited?	1 2 3 4 5 6 7 Images / Mixed / Location

Participant Instructions: Cued Recall and Subjective Ratings

You will be asked to answer five questions about details within the video clip (e.g., what colour was the dog?). Additionally, you will be asked to subjectively rate your memory for each video clip on several properties:

1. Emotional Intensity: Emotions can be felt with different intensities. For example, a negative memory can be extremely sad while another may be just somewhat sad. Similarly, a positive memory can be really exciting and euphoric and another can be calm and relaxing. Regardless of whether the memory was positive or negative, how intense was the emotion?

As I remember the (video clip), the emotional intensity I feel is:

1 2 3 4 5 6 7

None Minimal Moderate High

2. Reliving: This question relates to a feeling of experiencing the event again as if it were happening right now, or as if you were mentally traveling back in time to when the event occurred. How much do you feel that you can relive the memory?

As I remember the (video clip), the degree of reliving I feel is:

1 2 3 4 5 6 7

None Minimal Moderate High

3. Vividness: This question refers to the clarity with which you can see the event in your mind. The visual resolution of some memories may be as clear as if watching a high-definition show in front of you, whereas the visual resolution of other memories may be much poorer.

As I remember the (video clip), the vividness of my recollection is:

1 2 3 4 5 6 7

None Minimal Moderate High

4. Belief in memory accuracy: This question refers to how strongly you believe that your memory of the event in the video clip is an accurate representation of what actually happened. We may be highly confident in the accuracy of some memories, while not at all confident in the accuracy of others.

As I remember the (video clip), I feel the accuracy of my memory is:

1	2	3	4	5	6	7
None		Minimal		Moderate		High

Univariate Analysis of the Encoding and Retrieval Sessions. To confirm that our experimental design activated expected memory encoding and retrieval regions, regardless of type of visuotactile congruence, we conducted univariate analyses for each fMRI session separately. These analyses were a sanity check to ensure that our experimental design activated brain regions commonly implicated in each phase of memory processing. We used the same trial-wise beta-estimates from the encoding and retrieval sessions as in the multivariate analyses, except that they were spatially smoothed with a 8mm full-width at half maximum Gaussian kernel consistent with standards for univariate analyses (Mikl et al., 2008). For the encoding session, beta weights reflecting the effects of forming memories for each condition (i.e., average effect of the synchronous + asynchronous condition) taken from the last 20 seconds of each trial when participants were instructed to remember as much as possible about each scene were contrasted against baseline estimates for each participant. The first 20 seconds of each trial during the encoding session involved synchronous or asynchronous visuotactile stimulation of the mannequin's body against a still-frame image of the opening shot of each video to either induce a sense of ownership over the mannequin in the synchronous condition or reduce ownership over the mannequin in the asynchronous condition. These first 20 seconds were omitted from the analysis as they did not involve a memory task. For the retrieval session, beta weights that reflected the 17.5 seconds of memory retrieval (i.e., from the "Close Eyes" cue, see Figure 2B) were contrasted against baseline estimates for each participant. We visualized activations present at the voxelwise uncorrected threshold of $p < .001$ for each session, as the aim of these univariate analyses was to verify that well-known memory areas are active during encoding and retrieval in a purely descriptive approach.

Supplementary Results

Presence. Ratings from the three presence statements were averaged together separately for the synchronous (M = 3.38, SD = 1.14) and asynchronous (M = 3.64, SD = 1.29) visuotactile congruence conditions. A paired sample t-test did not reveal any significant differences between conditions, $F(1,29) = 1.32$, $p = .26$, Cohen's $d = .04$, demonstrating that both types of visuotactile stimulation conditions induced similar levels of presence within the immersive videos.[1]

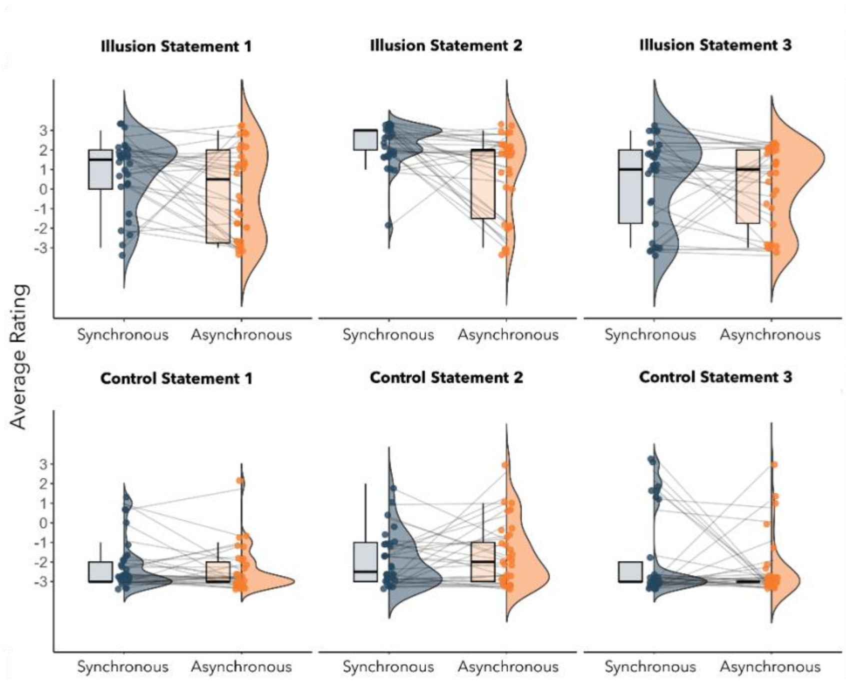
Univariate Analysis of the Encoding and Retrieval Sessions. We observed activation of many regions involved in encoding the immersive pre-recorded videos, regardless of the synchronicity of visuotactile stimulation (see Supplementary Figure 5). Large clusters of activity in the bilateral superior temporal gyri extended medially into both hippocampi, rostrally to dorsolateral prefrontal cortex, and caudally to posterior visual regions. We also identified activation in the bilateral dorsal posterior cingulate cortex and superior parietal gyrus. These findings are consistent with the previous memory encoding literature and confirms that participants were encoding the events as instructed (Gottlieb,

[1] Restricting the analysis to only those participants included in the fMRI analyses (N = 24) led to the same pattern of results ($t(23) = 1.33$, $p = .20$).

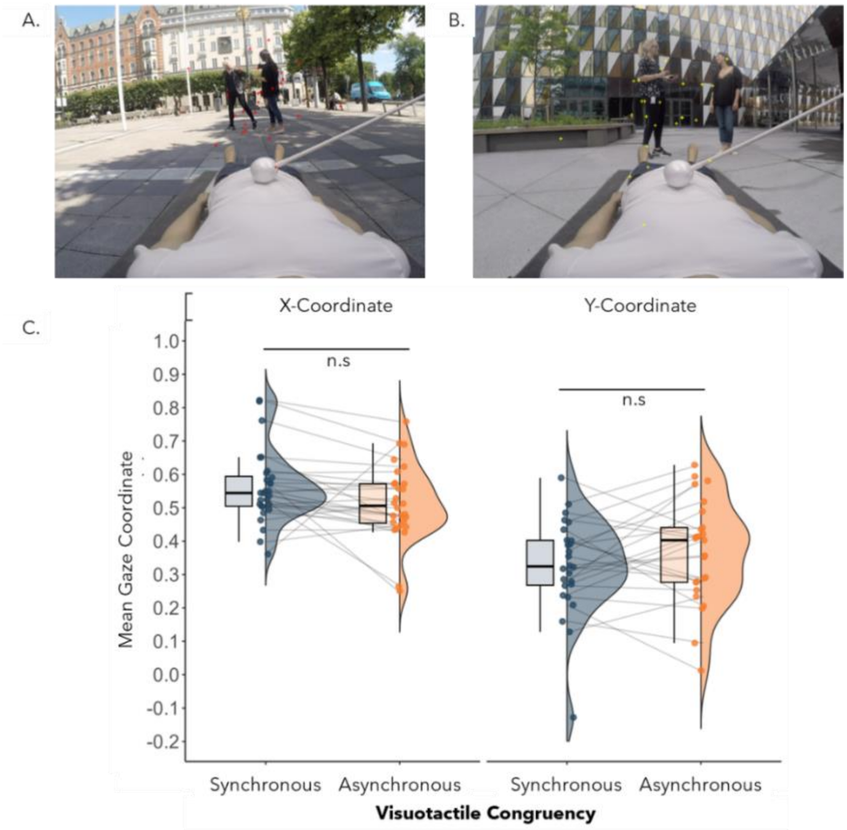
Wong, de Chastelain & Rugg, 2012; Kim, 2015, 2019; Sonkusare et al., 2019; Spaniol et al., 2009; Rugg et al., 2015).

As expected, the retrieval of the immersive videos from memory recruited a more restricted, lateralized network of brain regions which included the left dorsomedial prefrontal cortex, dorsolateral prefrontal cortex, precuneus, and angular gyrus (Supplementary Figure 6). These regions are typically reported in the memory literature (Iriye & St. Jacques, 2019; Svoboda et al., 2006), and confirm that participants in our study were retrieving memories for the immersive videos as instructed.

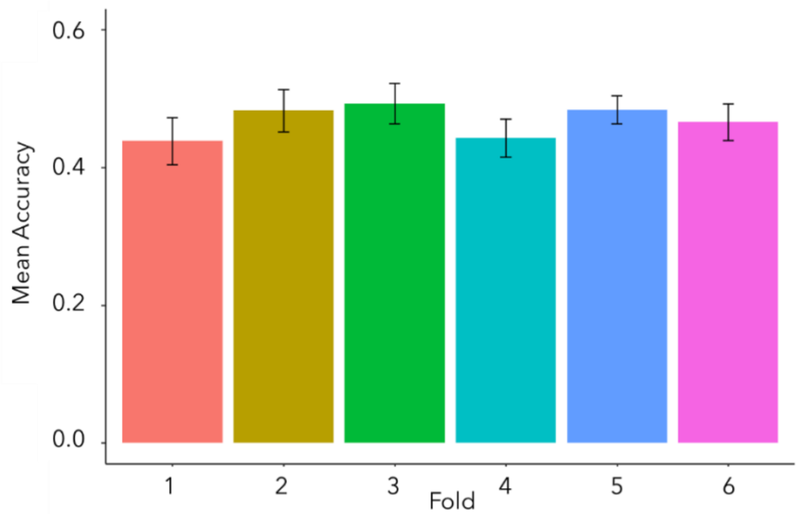
Supplementary Figures



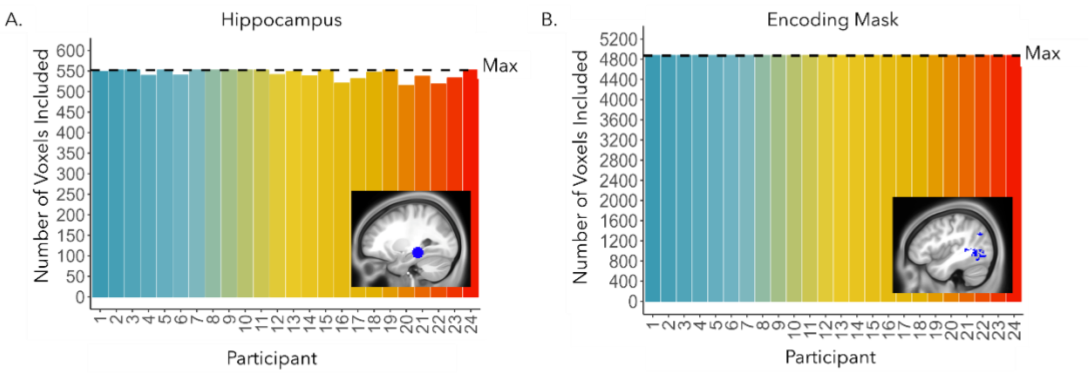
Supplementary Figure 1. Individual responses to each questionnaire statement item are depicted.



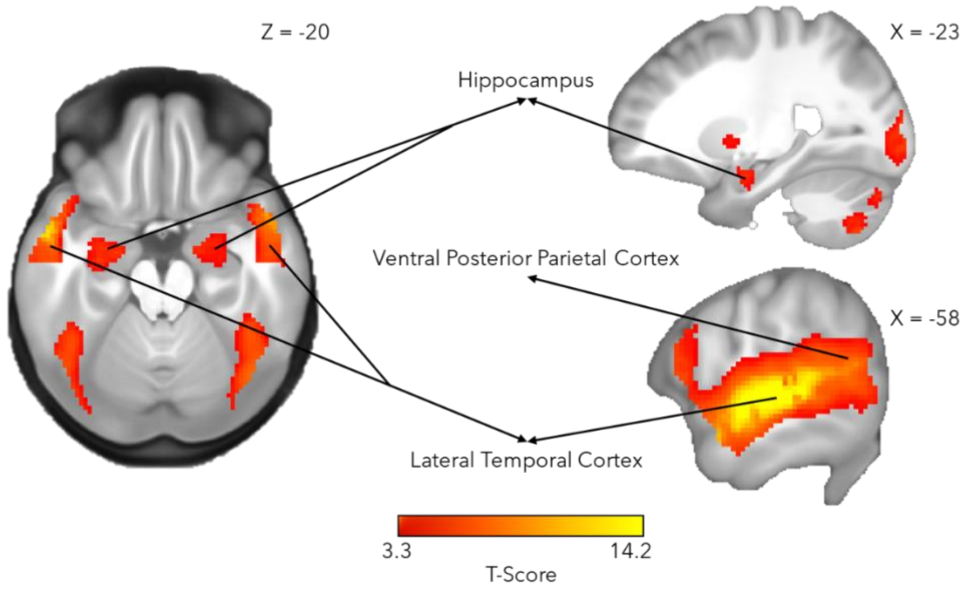
Supplementary Figure 2. Average fixation coordinates did not differ between the synchronous (A) and asynchronous (B) conditions. Each cross represents the average fixation across the video for one participant. There was no main effect of type of visuotactile stimulation on mean gaze (p=0.22), or interaction between type of visuotactile stimulation and co-ordinate (p = .19; C).



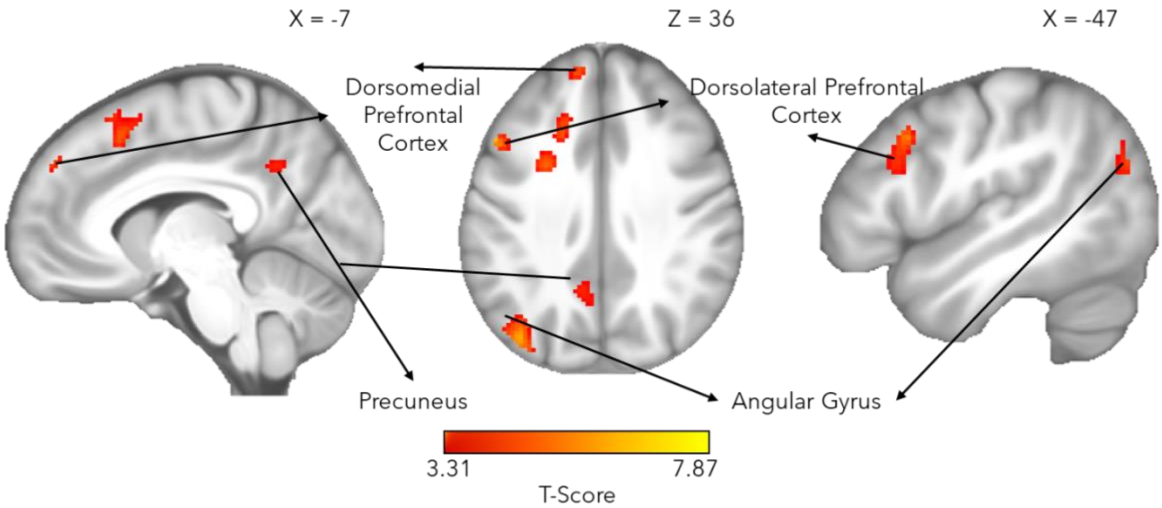
Supplementary Figure 3. Average classification accuracy from the memory retrieval session for each fold of the leave-one-out crossvalidation procedure in the hippocampal ROI analysis is depicted. Classifier accuracy was inconsistent across folds, with folds one and four performing worse than other folds.



Supplemental Figure 4. (C) Out of a total 552 possible voxels present in the hippocampal ROI, neural data from an average of 542.83 (*SD* = 11.87) across subjects were extracted for the RSA analysis (i.e., 98.34%). The dotted line indicates the maximum possible number of voxels within the ROI. (D) Neural data from all 4874 possible voxels in the encoding mask ROI were extracted for the RSA analysis. The dotted line indicates the maximum possible number of voxels within the ROI.



Supplemental Figure 5. *Brain Regions Involved in Memory Encoding*. Forming memories for the immersive videos activated brain regions such as bilateral hippocampus, lateral temporal cortex, ventral posterior parietal cortex, and visual cortex. *N* = 24. . For descriptive purposes, the statistical map was thresholded at $p < .001$ (uncorrected) and overlaid onto the average T1w images of 440 subjects obtained from the WU-Minn HCP dataset (van Essen et al., 2013).



Supplemental Figure 6. *Brain Regions Involved in Memory Retrieval*. Retrieving memories for the immersive videos activated brain regions such as the left dorsomedial prefrontal cortex, dorsolateral prefrontal cortex, precuneus, and angular gyrus (N = 24). For descriptive purposes, the statistical map was thresholded at $p < .001$ (uncorrected) and overlaid onto the average T1w images of 440 subjects obtained from the WU-Minn HCP dataset (van Essen et al., 2013).

Supplemental Discussion

To verify that participants were forming and recalling memories as expected during the experiment and support the interpretation of our main multivariate effects, we conducted two univariate general linear model analyses on the encoding and retrieval sessions separately. Encoding the immersive videos into memory in the present study was linked to extensive bilateral activation of the superior and middle temporal gyri, hippocampus, lateral frontal cortex, superior and inferior parietal cortex, and posterior visual cortex. Our findings are consistent with the idea that the specific brain regions involved in encoding are largely unspecialized and highly contingent upon the demands imposed by the task in question (Rugg et al., 2015). Specifically, the visual aspect of the video stimuli can be traced to activation of the posterior visual and inferior temporal regions we observed (Gottlieb et al., 2012; Park & Rugg, 2011; Sonkusare et al., 2019). The activation of the superior temporal cortex was likely related to the processing of auditory information contained in each video (Gottlieb et al., 2012; Park & Rugg, 2011; Sonkusare et al., 2019), and/or the processing of social interactions between the individuals in the scene (Lahnakoski et al., 2012). The lateral frontal and parietal regions identified in our analysis are part of the frontoparietal cognitive control network, which mediates attentional mechanisms required to successfully process and store task-relevant information in long-term memory (Jablonowski & Rose, 2022; Kim, 2015; Thomas Yeo et al., 2011). Activation of the hippocampus is frequently reported during successful encoding and is thought to keep a record of the patterns of cortical activity present as an event unfolds so that those patterns can be reinstated during memory retrieval (Kim, 2015, 2019; Rugg et al., 2015; Simons et al., 2022). The results of subsequent memory paradigms, which contrast activity linked to encoding later remembered compared to later forgotten information, confirm that the brain regions we observed as active during encoding are implicated in successful memory formation (Kim, 2015; Spaniol et al., 2009).

In contrast to the highly distributed number of regions activated at encoding, retrieval of the immersive videos activated a more restricted network of brain regions, including the left medial prefrontal cortex, left dorsolateral prefrontal cortex, left posterior parietal regions including the angular gyrus, and the left precuneus. These regions are typically reported in the memory literature (Iriye & St. Jacques, 2019; Svoboda et al., 2006), and confirm that participants in our study were retrieving memories for the immersive videos as instructed. The dorsomedial prefrontal cortex has been implicated in self-related processing during autobiographical memory retrieval (Svoboda, McKinnon, & Levine, 2006). Activation of frontal and parietal regions during retrieval has been linked to control processes required to retrieve semantic information associated with the memory (Binder & Desai, 2011; Dosenbach et al., 2007; Lepage et al., 1999.; Vincent et al., 2008). The precuneus is involved in mental imagery during autobiographical retrieval (Byrne et al., 2007; Cabeza & St Jacques, 2007; Cavanna & Trimble, 2006), particularly mental imagery related to adopting a specific egocentric perspective within a remembered scene (Iriye & St. Jacques, 2020; St. Jacques et al., 2017).

In sum, the results of the univariate analyses confirm that participants activated brain regions typically associated with the formation of memories for multisensory stimuli as they encoded the immersive videos in the present study, and regions linked with remembering past events when they retrieved these videos from memory. Thus, we can be confident that the multivariate effects of body ownership on encoding and reinstatement we observed in the present study reflect genuine memory processes. In turn, the lack of significant decoding accuracy in the retrieval session according to synchronicity of visuotactile stimulation at encoding is not due to an inability of participants to retrieve memories for the immersive videos.

References

Binder, J. R., & Desai, R. H. The neurobiology of semantic memory. In Trends in Cognitive Sciences. 2015: 15(11):527–536.

Byrne, P., Becker, S., & Burgess, N. Remembering the past and imagining the future: a neural model of spatial memory and imagery. Psychological Review. 2007:114(2):340–375.

Cabeza, R., & St Jacques, P. Functional neuroimaging of autobiographical memory. In Trends in Cognitive Sciences. 2007: 11(5):219–227.

Cavanna, A. E., & Trimble, M. R. The precuneus: A review of its functional anatomy and behavioural correlates. In Brain. Oxford University Press. 2006:129(3): 564–583.

Dosenbach, N. U., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., T Dosenbach, R. A., Fox, M. D., Snyder, A. Z., Vincent, J. L., Raichle, M. E., Schlaggar, B. L., & Petersen, S. E. Distinct brain networks for adaptive and stable task control in humans. 2007: 104(26):11073-11078.

Gottlieb, L. J., Wong, J., de Chastelaine, M., & Rugg, M. D. Neural correlates of the encoding of multimodal contextual features. Learning and Memory. 2012:19(12): 605–614.

Iriye, H., & St. Jacques, P. L. How visual perspective influences the spatiotemporal dynamics of autobiographical memory retrieval. Cortex. 2020:129:464–475.

Jablonowski, J., & Rose, M. The functional dissociation of posterior parietal regions during multimodal memory formation. Human Brain Mapping. 2022:43(11):3469-3485.

Kim, H. Neural correlates of explicit and implicit memory at encoding and retrieval: A unified framework and meta-analysis of functional neuroimaging studies. Biological Psychology. 2019:145(2):96–111.

Kim, H. Encoding and retrieval along the long axis of the hippocampus and their relationships with dorsal attention and default mode networks: The HERNET model. Hippocampus. 2015:25(4):500–510.

Lahnakoski, J. M., Glerean, E., Salmi, J., Jääskeläinen, I. P., Sams, M., Hari, R., & Nummenmaa, L. Naturalistic fMRI mapping reveals superior temporal sulcus as the hub for the distributed brain network for social perception. Frontiers in Human Neuroscience, 2012:6:233

Lepage, M., Ghaffar, O., Nyberg, L., & Tulving, E. Prefrontal cortex and episodic memory retrieval mode. *Proceedings of the National Academy of Sciences*. 2000:*97*(1):506-511.

Mikl, M., Mareček, R., Hluštík, P., Pavlicová, M., Drastich, A., Chlebus, P., Brázdil, M., & Krupa, P. Effects of spatial smoothing on fMRI group inferences. Magnetic Resonance Imaging, 2008:26(4):490–503.

Park, H., & Rugg, M. D. Neural correlates of encoding within-and across-domain inter-item associations. *Journal of cognitive neuroscience*. 2011:23(9):2533-2543.

Rugg, M. D., Johnson, J. D., & Uncapher, M. R. Encoding and retrieval in episodic memory: Insights from fMRI. *The Wiley handbook on the cognitive neuroscience of memory*. 2015: 84-107.

Simons, J. S., Ritchey, M., & Fernyhough, C. Brain Mechanisms Underlying the Subjective Experience of Remembering. Annual Review of Psychology. 2022:73: 159–186.

Sonkusare, S., Breakspear, M., & Guo, C. Naturalistic Stimuli in Neuroscience: Critically Acclaimed. In Trends in Cognitive Sciences. 2019:23(8): 699–714.

Spaniol, J., Davidson, P. S. R., Kim, A. S. N., Han, H., Moscovitch, M., & Grady, C. L. Event-related fMRI studies of episodic encoding and retrieval: Meta-analyses using activation likelihood estimation. Neuropsychologia. 2009:47(8–9):1765–1779.

St. Jacques, P. L., Szpunar, K. K., & Schacter, D. L. Shifting visual perspective during retrieval shapes autobiographical memories. NeuroImage. 2017:148:103–114.

Svoboda, E., McKinnon, M. C., & Levine, B. The functional neuroanatomy of autobiographical memory: A meta-analysis. Neuropsychologia. 2006: 44(12):2189–2208.

Thomas Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fisch, B., Liu, H., & Buckner, R. L. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of Neurophysiology. 2011:106(3): 1125–1165.

van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K. The WU-Minn Human Connectome Project: An overview. NeuroImage. 2013:80: 62–79.

Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., & Buckner, R. L. Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. Journal of Neurophysiology. 2008:100(6):3328–3342.