SUPPLEMENTARY INFORMATION

Self-judgment on emotional items modulates memory resources in breast cancer patients: a longitudinal MRI study

Joue et al.

S1 Demographics

Table S1 summarizes participant demographics and the neuropsychological scores pertinent to this paper, specifically tests for episodic memory, self representation, anxiety and depression, as well as general cognition and attention.

S2 Methods

S2.1 Neuropsychological battery

Participants completed a neuropsychology test battery including the French version of the Questionnaire of Self-Representation (QSR)[1], consisting of 50 positive and negative self descriptions dealing with different categories of self representation (including physical, moral, ethical, personal, familial, and social) and requiring responses on a 4-point Likert scale. Subscore scales include structural self or stability of self representation (QSR.certitude, `QSR.cert`) and functional self or self-esteem (QSR.esteem), both calculated on the basis of all 50 statements (i.e. across all self categories) and expressed as percentages. QSR.certitude is an index of the stability of self-knowledge as a trait and reflects the number of clear and consistent self-representation responses (i.e. 1/"Does not describe me at all" and 4/"Describes me absolutely"; the other points on the scale, 2/"Describes me a little" and 3/"Describes me well" are treated as vague responses). The higher the QSR.certitude score, the more certain the self-representation is. QSR.esteem measures self-esteem (also referred to as "valence"[2]). The higher the QSR.esteem score, the more positive the self-representation.

Anxiety was assessed with the State-Trait Anxiety Inventory (STAI)[3]. State anxiety (STAI-S) is a measure of situational anxiety, with participants being asked to respond based on "how you feel right now" (corresponding here to the period of breast cancer diagnosis for patients). Trait anxiety (STAI-T) is a measure of the general tendency to be anxious, with questions phrased by, "how you generally feel". Each subscale consists of 20 items scored on a four-point Likert-type scale. Subscale scores range from 20 to 80, with higher scores indicating increased anxiety.

Episodic memory was tested verbally (list of 16 words) and visually (set of eight abstract symbols) with the Encoding, Storage, Retrieval (ESR) paradigm[4–6]. Recognition performance for superficially and incidentally encoded items was used to assess episodic memory encoding ability. Free recall performance of deeply and intentionally encoded items was used as a measure of episodic memory retrieval.

The timed, paper-based Trail Making Test (TMT) was used to measure attention and cognitive flexibility. Both parts were administered: Part A (connect 25 circles of numbers randomly distributed on page in ascending order) and Part B (connect 25 circles with 1–13 numbers, letters A–L, alternating between numbers and letters). The completion time difference between parts B and A is used as an indicator of executive function[7].

S2.2 Behavioral models

Interpretations of model effect sizes follow Cohen's recommendations[8]. Linear mixed models fitted to test the predictability of the difference in memory for self-reference (`SR`) over semantic judgment of the emotional valence of a character trait (`SEM`) by group, testing time point, item emotional valence, QSR certitude score, and STAI trait anxiety score (full equation in Sec. S3.1.1) moderately explained the variance in memory performance based on the fixed effects alone (marginal $R^2 = 0.17$) and had substantial total explanatory power when random effects allowing emotional valence to vary by individual are taken into account (conditional $R^2 = 0.37$; see Table S2 for estimates). The overall effect sizes of group or treatment differences in SRE were negligible. The most influential predictors of the magnitude of SRE were STAI.T and QSR.certitude scores combined, which had different predicted effects in HC or PT depending on testing time point (colon indicates interaction, `grp:timept:QSR.cert:STAI.T`; Fig. S2) and the emotional valence of the memory item (`grp:timept:emo:QSR.cert`; `grp:timept:emo:STAI.T`).

The random term of the linear mixed models was determined using likelihood ratio tests on null models as well as comparing AICs of nested terms in models that were otherwise specified with maximal fixed terms (base R and R package *MuMIn*). Results from both suggested that the random effects term best describing the data was random intercept by individual and

Table S1. Demographics and neuropsychological test score averages ± standard deviations for study healthy participants (HC) and breast cancer patients (PT).

	HC		PT							
cancer grade	(N = 20)		all (N = 20)		*I* (N = 5)		*II* (N = 8)		*III* (N = 7)	
age [yrs]	55.6 ± 7.3		54 ± 4.7		53.4		53.2		55.1	
education [yrs]	12.1		11.4		10		11.1		12.9	
timept	t0m	t18m	t0m	t18m	t0m	t18m	t0m	t18m	t0m	t18m
Anxiety and depression										
STAI-S$^{\tau}$	28.2 ± 5.6	12.1 ± 3	31.8 ± 8.9	11.4 ± 3.3	27.4 ± 8.4	10 ± 2.5	31.6 ± 7.2	11.1 ± 3.6	35.1 ± 10.6	12.9 ± 3.4
STAI-T$^{\tau}$	37.2 ± 9.8	27.5 ± 5.5	36 ± 9.3	27.3 ± 9.3	33.8 ± 11.2	23.6 ± 4.6	36 ± 6.5	24.4 ± 4.8	37.4 ± 11.5	33.3 ± 12.9
BDI-SF	1.4 ± 1.4	–	3 ± 3.7	3 ± 2.4	3.4 ± 4.9	2.8 ± 1.3	2.4 ± 3.3	2.6 ± 2.3	3.4 ± 3.8	3.6 ± 3.3
General cognition										
MMSE	29.2 ± 1	29.1 ± 0.9	28.8 ± 1.1	28.5 ± 1.4	28.4 ± 0.9	28.2 ± 1.1	28.9 ± 1	28.4 ± 1.8	28.9 ± 1.5	28.9 ± 1.3
Verbal episodic memory: BEM										
Recall Superf	5.6 ± 2.2	5.7 ± 2.1	5 ± 1.5	5.8 ± 1.8	5.8 ± 1.3	6.2 ± 1.6	4.2 ± 1.3	5 ± 1.1	5.4 ± 1.5	6.6 ± 2.2
Recall Deep	9.8 ± 2	10.7 ± 2.5	8.3 ± 2.6	9.2 ± 2	9.4 ± 1.5	8.8 ± 1.5	7.2 ± 2.9	9.4 ± 2.1	8.9 ± 2.5	9.4 ± 2.4
Recog Superf	13.6 ± 1.8	13.8 ± 1.4	12.9 ± 2.3	13.4 ± 2.3	14.4 ± 1.1	13.2 ± 2.6	11.6 ± 2.5	12.6 ± 2.3	13.3 ± 2.1	14.6 ± 1.7
Recog Deep	15.9 ± 0.2	15.9 ± 0.2	15.8 ± 0.4	15.8 ± 0.4	15.8 ± 0.4	15.8 ± 0.4	15.5 ± 0.5	15.8 ± 0.5	16 ± 0	16 ± 0
Visual episodic memory: BEM										
Recall Superf	6.2 ± 2.1	7.3 ± 1.8	7.4 ± 2.5	8.7 ± 2.5	8 ± 2.5	8.8 ± 3.3	6.1 ± 2	7.8 ± 1.8	8.4 ± 2.8	9.9 ± 2.7
Recall Deep	10.9 ± 2.6	10.1 ± 2.7	9.9 ± 2.9	10.4 ± 3	10 ± 2.3	10.2 ± 2.9	8.5 ± 2.7	9.9 ± 2.4	11.6 ± 3.1	11.1 ± 3.8
Recog Superf	7.3 ± 0.9	7.6 ± 0.6	7.7 ± 0.6	7.7 ± 0.6	7.4 ± 0.9	7.2 ± 1	7.6 ± 0.5	7.8 ± 0.5	7.9 ± 0.4	8 ± 0
Recog Deep	7.3 ± 0.9	7.8 ± 0.4	7.4 ± 0.8	7.5 ± 0.8	6.8 ± 1.3	7.5 ± 1	7.4 ± 0.5	7.1 ± 0.8	7.9 ± 0.4	7.9 ± 0.4
Executive functions: TMT										
TMT-A *t* [ms]	33.9 ± 9	36.6 ± 11.2	35.8 ± 12.1	32.4 ± 8	31.2 ± 5.7	32.6 ± 8.9	38.3 ± 16.1	33.5 ± 10.7	36.3 ± 10.7	31 ± 3.9
TMT-A # err‡	0 ± 0.2	0 ± 0	0.2 ± 0.4	0.2 ± 0.4	0.2 ± 0.4	0.4 ± 0.5	0.2 ± 0.5	0 ± 0	0.1 ± 0.4	0.3 ± 0.5
TMT-B *t* [ms]	74 ± 26.4	80.8 ± 27.7	80.7 ± 22.8	91.2 ± 34.8	82.6 ± 15.1	96.2 ± 28.9	81.1 ± 32.2	96.2 ± 49.1	78.7 ± 16.7	81.9 ± 17.9
TMT-B # err†	0.1 ± 0.5	0.3 ± 0.6	0.6 ± 0.7	0.6 ± 0.8	1 ± 0.7	0.6 ± 0.9	0.2 ± 0.5	0.8 ± 0.9	0.6 ± 0.8	0.3 ± 0.5
TMT-B # persev. err‡	0 ± 0	0 ± 0	0.2 ± 0.5	0.4 ± 0.6	0 ± 0	0.4 ± 0.5	0.1 ± 0.4	0.4 ± 0.7	0.4 ± 0.8	0.4 ± 0.5
TMT $t_B - t_A$ [ms]	40 ± 22	44.2 ± 21.5	44.8 ± 20.9	58.8 ± 32.4	51.4 ± 12.2	63.6 ± 27.7	42.8 ± 30.6	62.7 ± 45.7	42.4 ± 11.9	50.9 ± 16.4
Self representation: QSR										
Certitude	47 ± 21.3	46.9 ± 17.8	52.3 ± 17.8	54.1 ± 18.9	60.8 ± 23.4	61.6 ± 22.6	55 ± 15.4	53.8 ± 20.4	43.1 ± 13.8	49.1 ± 15.2
Self esteem	69.2 ± 7.5	71.2 ± 6.5	72 ± 9.5	73.6 ± 10.6	76.1 ± 12.7	77.3 ± 10.5	72.6 ± 8.3	75.2 ± 12.7	68.3 ± 8.2	69.2 ± 7.8

Tests (with theoretical score ranges): STAI-S (20–80), STAI-T (20–80); BDI-SF (0–39; 0–9 minimal depression); Beck's Depression Inventory, short form (BDI-SF); Mini-Mental State Exam (MMSE) (0–30); Verbal BEM Recall Superf (0–16), Recall Deep (3–14), Recog Superf (0–16), Recog Deep (0–16); Visual Recog Superf (0–16), Recog Deep (0–16), Recog Superf (0–8), Recog Deep (0–8); Working memory (WM) span (0–16) for category "place" forwards (fwd) and in reverse (rev); Trail Making Test (TMT) A and B speed (*t*; normal ranges, Part A: 1–39 s, Part B: 1–91 s), number of errors (# err), and number of persevering errors (# persev. err); Verbal fluency test (number of words within 60 s) for phonetic fluency with words beginning with "P") and semantic fluency with category "animal"; Questionnaire of Self-Representation (QSR).

† Group difference at t0m (Welch's 2-sample t-test, $p < 0.05$)

‡Group difference at t18m (Welch's 2-sample t-test, $p < 0.05$)

$^{\tau}$Difference between t0m and t18m in PT (Welch's 2-sample t-test, $p < 0.05$)

individual slopes by emotional valence. Contribution of fixed effect terms were assessed (R package *lmerTest*) using backward stepwise elimination of the term with the lowest F-stats/highest p-value for a given model, starting with $p < 0.1$ from a maximally specified model, until no terms with lower F-stats remain or no terms are left in the model?a procedure proposed by biostatisticians><>zuur.EA:09.

As STAI.T and STAI.S were highly correlated, as were QSR.certitude and QSR.esteem, only STAI.T and QSR.certitude were used in all models. QSR.certitude and STAI.T were anti-correlated in HC (t0m: Spearman's $\rho = -0.50$, 95% CI$[-0.77, -0.075]$, $p_{FDR} = 0.025$; t18m: $\rho = -0.57$, $[-0.81, -0.17]$, $p_{FDR} = 0.0084$), but only before chemotherapy in PT (t0m: $\rho = -0.63$, $[-0.84, -0.27]$, $p_{FDR} = 0.0024$; t18m: $\rho = -0.27$, $[-0.63, 0.20]$, $p_{FDR} = 0.26$), indicating that they were related but different measures. Their interactions were therefore tested in all models.

Visual diagnostics were done including checks for nonlinear interaction effects and over-extrapolation (R package *interflex*)[9]. Noted was the more limited range of STAI.T especially in HC at t18m. The final model was refitted (R package *lme4* estimated using REML and nloptwrap optimizer; interaction effects for plotting calculated with R package *ggeffects*). Confidence Intervals (95% CIs) and p-values were computed using a Wald t-distribution approximation.

According to distribution fit tests, the logarithmic transform was better than the inverse or gamma functions for the left-skewed distribution of response times (RTs) during the memory task, and therefore, all parametric statistical tests were performed on logarithmically transformed RTs.

S2.3 MRI methods

fMRIprep[10,11] (RRID:SCR_016216), based on *Nipype* 1.5.1 (RRID:SCR_002502)[12,13], was used to preprocess the data. The following boilerplate text of the preprocessing steps was automatically generated by fMRIPrep and released under the CC0 license.

S2.3.1 Anatomical data preprocessing

A total of 2 T1-weighted (T1w) images were found within the input BIDS dataset. All of them were corrected for intensity non-uniformity (INU) with `N4BiasFieldCorrection`[14], distributed with ANTs 2.3.3 [15, RRID:SCR_004757]. The T1w-reference was then skull-stripped with a *Nipype* implementation of the `antsBrainExtraction.sh` workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using `fast` (FSL 5.0.9, RRID:SCR_002823)[16]. A T1w-reference map was computed after registration of 2 T1w images (after INU-correction) using `mri_robust_template` (FreeSurfer 6.0.1)[17]. Brain surfaces were reconstructed using `recon-all` (FreeSurfer 6.0.1, RRID:SCR_001847)[18], and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentation of the cortical gray-matter of Mindboggle (RRID:SCR_002438)[19]. Volume-based spatial normalization to two standard spaces (MNI152NLin2009cAsym, MNI152NLin6Asym) was performed through nonlinear registration with `antsRegistration` (ANTs 2.3.3), using brain-extracted versions of both T1w reference and the T1w template. The following templates were selected for spatial normalization: *ICBM 152 Nonlinear Asymmetrical template version 2009c* (RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym)[20], *FSL's MNI ICBM 152 non-linear 6th Generation Asymmetric Average Brain Stereotaxic Registration Model* (RRID:SCR_002823; TemplateFlow ID: MNI152NLin6Asym)[21].

S2.3.2 Functional data preprocessing

For each of the 10 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Susceptibility distortion correction (SDC) was omitted. The BOLD reference was then co-registered to the T1w reference using `bbregister` (FreeSurfer) which implements boundary-based registration[22]. Co-registration was configured with six degrees of freedom. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using `mcflirt` (FSL 5.0.9)[23]. BOLD runs were slice-time corrected using `3dTshift` from AFNI 20160207 (RRID:SCR_005927)[24]. The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying the transforms to correct for head-motion. These resampled BOLD time-series will be referred to as *preprocessed BOLD in original space*, or just *preprocessed BOLD*. The BOLD time-series were resampled into standard space, generating a *preprocessed BOLD run in MNI152NLin2009cAsym space*. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Automatic removal of motion artifacts using independent component analysis (ICA-AROMA)[25] was performed on the *preprocessed BOLD on MNI space* time-series after removal of non-steady state volumes and spatial smoothing with an isotropic, Gaussian kernel of 6mm FWHM (full-width half-maximum). Corresponding "non-aggressively" denoised runs were produced after such smoothing. Additionally, the "aggressive" noise-regressors were collected and placed in the corresponding confounds file. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power

(absolute sum of relative motions)[26] and Jenkinson (relative root mean square displacement between affines[23]). FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* following the definitions by Power and colleagues[26]. The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (*CompCor*)[27]. Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, the aCompCor masks are subtracted a mask of pixels that likely contain a volume fraction of GM. This mask is obtained by dilating a GM mask extracted from the FreeSurfer's *aseg* segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each[28]. Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using `antsApplyTransforms` (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels[29]. Non-gridded (surface) resamplings were performed using `mri_vol2surf` (FreeSurfer).

Many internal operations of *fMRIPrep* use *Nilearn* 0.6.2 (RRID:SCR_001362)[30], mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in *fMRIPrep*'s documentation.

S2.3.3 Automatic structural segmentation

The Freesurfer longitudinal subroutine first creates an unbiased within-subject template using robust, inverse consistent registration of the structural images at both time points. This template served to initiate the parameters of the iterative, non-linear optimization computations involved in preprocessing (skull stripping, normalization to standard space, and atlas registration), volume segmentation, and surface reconstruction. Combined with probabilistic methods, the confounding effects of intra-individual structural differences are minimized for more robust and consistent estimations. Parcellation into cortical and subcortical regions using standardized atlases is done by following intensity gradients across tissue classes. For surface-based metrics like curvature, folding index, and sulcal depth, the surfaces of cortical models are inflated, registered to a spherical atlas using individual cortical folding patterns, and then parcellated using gyral and sulcal structure. Surface-based volume estimations was based on sulcal structure. The Destrieux Atlas of 148 cortical regions[31] and the 37-region FreeSurfer subcortical atlas aseg[32]) were chosen here.

S2.3.4 Data checks

Image data and preprocessing quality were checked visually through reports generated by fMRIPrep, MRIQC[33], and manually in image viewers at key check points (skull stripping, normalization, image co/registration). MRIQC produces interactive individual and group reports visually presenting various quality metrics that facilitate checks for blurring and ghosting caused by motion and artifacts, for example, in structural images, and temporospatial information (such as temporal signal to noise ratio) and artifacts in functional images.

fMRI modeling was checked for quality in several ways. For each contrast of interest for each participant, histograms of contrast estimate values were visually checked to be normally distributed about a mean of zero, indicating overall whole brain activation/deactivation and on the same scale across participants. Contrast estimates were also plotted against the square root of the residual mean squares, the estimated noise SD used to calculate the t-statistic in Gaussian random field data, to check that variances did not vary widely and were relatively heteroscedastic.

S2.3.5 Clustering analyses: UMAPs

All feature sets were symmetrized percent changes of structural and functional MRI measures, and of STAI.S, STAI.T, QSR.certitude, and QSR.selfEsteem scores, and completion time differences between TMT Parts A and B. Feature sets differed in sMRI and fMRI measures included. The smallest feature set (Fig. S3a, Hc.SR) included volumetric changes in the hippocampus (Hc), whole and anterior/posterior divisions, and beta estimates from the mass-univariate GLM of BOLD activation and gPPI estimates for `SR.H.pos>SR.H.neg` summarized as the eigenvalue of voxels within target atlas-defined

ROIs (including subregions) in the left amygdala (L-Amyg), left hippocampus (L-Hc), left thalamus (L-Th), medial prefrontal cortex (mPFC), left middle and superior temporal gyri (L-MTG, L-STG). The next bigger feature set (Fig. S3b, HcAmyTh.SR) also included volumetrics for the amygdala and thalamic divisions. The feature set next in size (Fig. S3c) also included volumetrics and surface-based metrics from all the regions from the standard atlases used and whole-brain measures, as well as beta estimates/gPPI estimates for additional contrasts: `SR.H.pos>OR.H.pos` and `SR.H.neg>OR.H.neg`. A set was also tested that additionally included beta and gPPI estimates for `SR` contrasted against `SEM` (Fig. S3d). The biggest feature set (Fig. S3e, all) was the previous feature set plus all the volumetrics and surface-based metrics provided by the atlases used.

As previous large-scale studies have found that older women and women who only had lumpectomies tended to be more positive[34], we also show the distribution of participants by cancer grade and surgery undergone in the clustering results.

S3 Results

S3.1 Behavioral

All participants performed above chance level in general (one-sided t-test against chance performance level of 50%; HC.t0m: $t_{19} = 0.63, p < 0.001$, Cohen's $d = 1.15$; HC.t18m: $t_{19} = 0.63, p < 0.001$, $d = 1.10$; PT.t0m: $t_{19} = 0.64, p < 0.001$, $d = 1.00$; PT.t18m: $t_{19} = 0.66, p < 0.001$, $d = 1.27$). Overall, PT tended to perform slightly worse than HC at t18m (Welch's 2-sample one-sided t-tests; t0m: $t_{35.4} = 0.14, p = 0.56$, $d = 0.044$; t18m: $t_{37.9} = -0.62, p = 0.27$, $d = -0.19$).

S3.1.1 Self-reference Effect (SRE)

The winning model for SRE was the following (colon indicates interaction):

$$\Delta\texttt{gmean}_{\texttt{SR.pos-OR.pos}} \sim \texttt{grp + timept + emo + QSR.cert + STAI.T +}$$
$$\texttt{grp:timept + grp:emo + timept:emo + grp:QSR.cert +}$$
$$\texttt{timept:QSR.cert + emo:QSR.cert +}$$
$$\texttt{grp:STAI.T + timept:STAI.T + emo:STAI.T + QSR.cert:STAI.T +}$$
$$\texttt{grp:timept:emo + grp:timept:QSR.cert + grp:emo:QSR.cert +}$$
$$\texttt{timept:emo:QSR.cert + grp:timept:STAI.T +}$$
$$\texttt{grp:emo:STAI.T + timept:emo:STAI.T +}$$
$$\texttt{grp:QSR.cert:STAI.T + timept:QSR.cert:STAI.T +}$$
$$\texttt{grp:timept:emo:QSR.cert + grp:timept:emo:STAI.T +}$$
$$\texttt{grp:timept:QSR.cert:STAI.T +}$$
$$\texttt{(1 + emo | subjID)}$$

The general self-reference effect (SRE) by group, emotional valence and time is shown in Fig. S1. Results from the best-fitting linear mixed model are shown in Fig. S2.

The winning linear mixed model for the difference in memory performance between positive and negative self-judged items was

$$\Delta\texttt{gmean}_{\texttt{SR.pos-SR.neg}} \sim \texttt{grp + timept + QSR.cert + STAI.T +}$$
$$\texttt{grp:timept + grp:QSR.cert + timept:QSR.cert + grp:STAI.T +}$$
$$\texttt{timept:STAI.T + QSR.cert:STAI.T +}$$
$$\texttt{grp:timept:QSR.cert + grp:timept:STAI.T + (1 | subjID)}$$

The winning linear mixed model for the difference in memory performance between positive self-judged items and positive other-judged items (asterisks indicate full interaction and lower-order terms) was

$$\Delta\texttt{gmean}_{\texttt{SR.pos-OR.pos}} \sim \texttt{grp * timept * QSR.cert * STAI.T + (1 | subjID)}$$

S3.1.2 Other-reference (OR)

There was no memory advantage for trait items judged of others (HC: all $p > 0.30$, PT: all $p > 0.80$). Neither group showed negativity or positivity when judging others (`OR.neg` and `OR.pos`: all $t_{19} < -0.8, p > 0.79$ except for `OR.neg` HC.t0m: $t_{19} = 0.85, p = 0.20, d = 0.19$; HC.t18m: $t_{19} = 1.3, p = 0.010, d = 0.29$). There were no overall group differences in positive or negative items when self-judged compared to judged of others (largest t-statistic magnitude for `neg`.t18m: $t_{39} = -1.1, p_{FDR} = 0.74$).

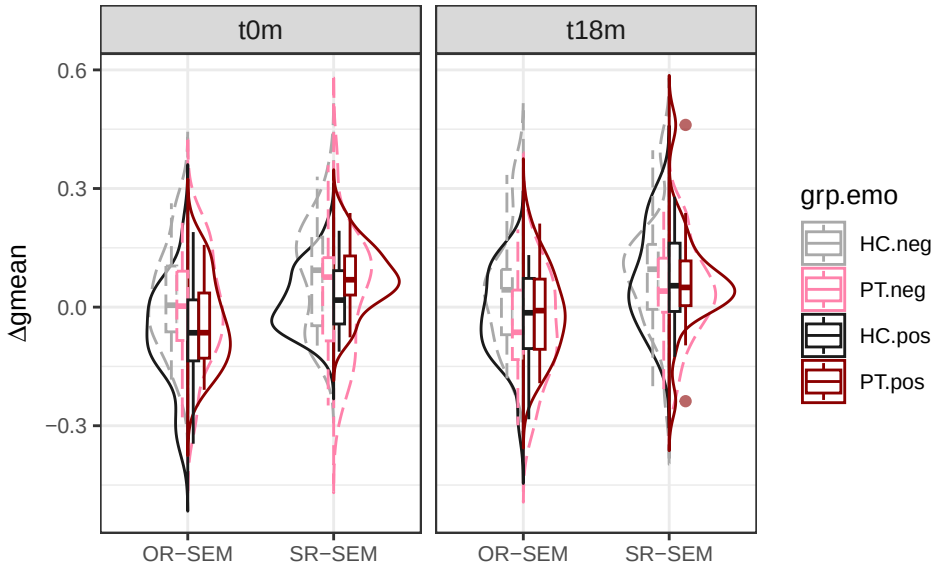
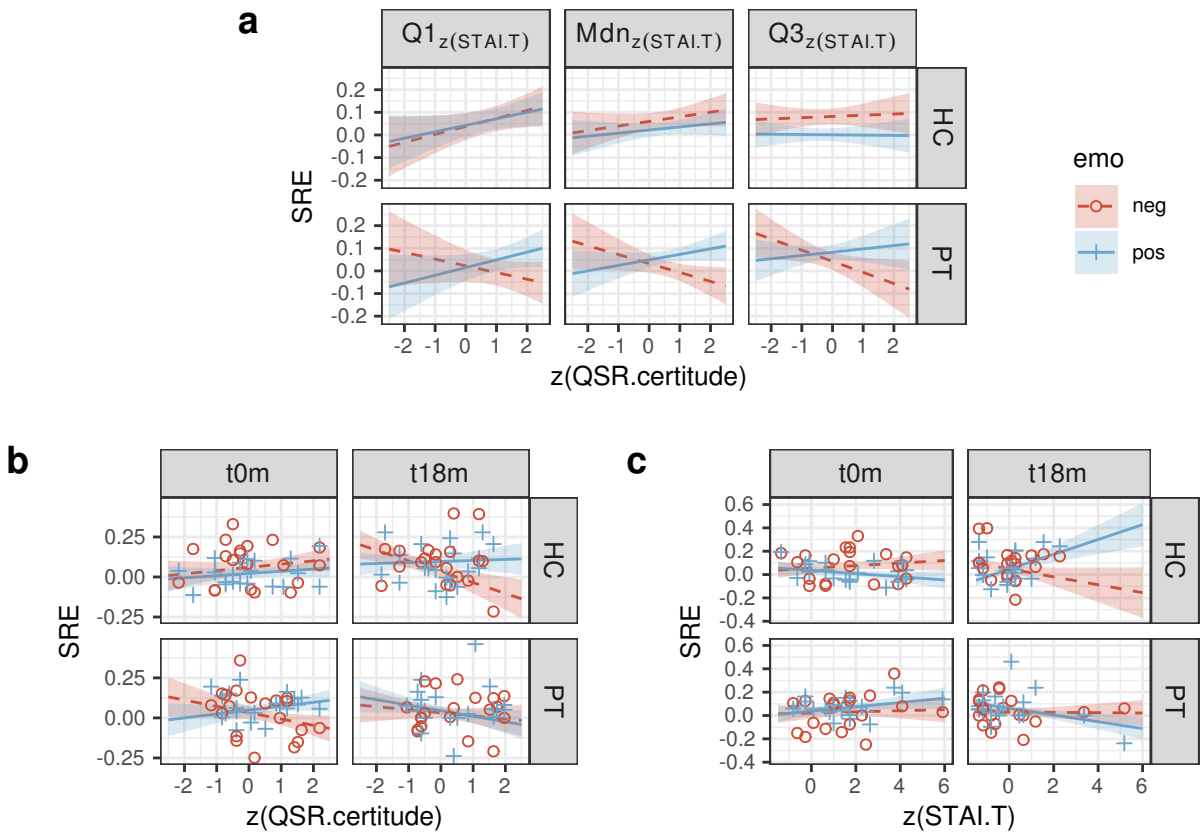


Figure S1. Differences in memory performance measured as the geometric mean (Δgmean) between self (SR) and other judgment (OR) compared to semantic questions (SEM) in the two groups (grp): healthy control (HC, shown with grayscale lines) and patient (PT, red-toned lines) groups at each testing time point (t0m: before chemotherapy in PT, in light colors; in t18m: 18 months from the start of chemotherapy in PT, in dark colors) for negative (neg; dashed lines) and positive (pos; solid lines) emotionally valenced memory items (emo). Both groups showed the self-reference effect (SRE $:= \Delta\text{gmean}_{SR-SEM} > 0$) at both testing times. There was no group difference in SRE. There was no memory advantage for items judged about another ($\Delta\text{gmean}_{OR-SEM} \approx 0$). Only HC showed a memory advantage for negatively valenced items. The geometric mean is calculated as the square root of the product of the true positive rate and the true negative rate of memory responses. A gmean of 0.50 indicates chance performance and a d(gmean) of 0 indicates no difference. Note density plots show memory items overlaid on top of each other by group, while boxplots are grouped by emotional valence. One-sided t-tests for overall SRE: HC.t0m $t_{39} = 2.8, p = 0.004, d = 0.45$; HC.t18m $t_{39} = 4.0, p < 0.001, d = 0.64$; PT.t0m $t_{39} = 2.7, p = 0.005, d = 0.43$; PT.t18m $t_{39} = 2.7, p = 0.005, d = 0.43$.



]

Figure S2. Predicted self-reference effect (SRE) based on linear mixed regression modeling. Linear mixed regression models indicated that there was a general SRE for negative memory items in HC, but that self-certitude (QSR.certitude) and trait anxiety (STAI.T) scores combined predicted SRE ($gmean_{SR} - gmean_{SEM}$) on the episodic memory task differently between healthy (HC) and patient (PT) groups (panel A) at the two testing time points (t0m and t18m) in and was modulated by the emotional valence (emo) of memory items (panels B and C). All scores (SRE, QSR.certitude, and STAI.T) have been scaled to the mean and standard deviation of respective scores of HC at t18m. Predictions shown in panel **a** are shown for the upper bound of the first quartile (Q1), median (Mdn) and third quartile (Q3) of scaled (z) STAI.T. Real data points are also plotted in panels **b** and **c**. In all panels, memory items with negative (neg) emotional valence are represented by red open circles and red dashed lines/shaded areas while those with positive (pos) emotional valence are blue crosses and blue solid lines/shaded areas. Shaded areas are standard errors of the mean.

Table S2. Estimates [with 95% confidence intervals] from mixed linear regression on memory performance differences for the self-reference effect SRE (`SR-SEM`), positive and negative SRE (`SRE.pos-SRE.neg`), and positive SRE vs. other-reference effect (`SRE.pos-ORE.pos`)

	estimates [95% CI]					
predictors	$\Delta gmean_{\texttt{SR-SEM}}$		$\Delta gmean_{\texttt{SRE.pos-SRE.neg}}$		$\Delta gmean_{\texttt{SRE.pos-ORE.pos}}$	
`grp[PT]`	−0.022	[−0.16, 0.12]	0.059	[−0.12, 0.24]	−0.0001	[−0.13, 0.13]
`timept[t18m]`	−0.001	[−0.099, 0.097]	0.040	[−0.096, 0.18]	−0.037	[−0.15, 0.073]
`emo[pos]`	−0.018	[−0.13, 0.093]				
`QSR.cert`	0.027	[−0.038, 0.092]	0.045	[−0.042, 0.13]	0.033	[−0.037, 0.10]
`STAI.T`	0.013	[−0.024, 0.049]	−0.009	[−0.058, 0.040]	−0.007	[−0.043, 0.028]
`grp[PT]:timept[t18m]`	0.003	[−0.14, 0.15]	−0.005	[−0.21, 0.20]	0.020	[−0.14, 0.18]
`grp[PT]:emo[pos]`	0.023	[−0.15, 0.19]				
`timept[t18m]:emo[pos]`	0.003	[−0.12, 0.13]				
`grp[PT]:QSR.cert`	−0.062	[−0.17, 0.040]	0.044	[−0.082, 0.17]	−0.013	[−0.12, 0.093]
`timept[t18m]:QSR.cert`	−0.031	[−0.12, 0.056]	0.036	[−0.087, 0.16]	−0.038	[−0.14, 0.062]
`emo[pos]:QSR.cert`	−0.007	[−0.079, 0.065]				
`grp[PT]:STAI.T`	−0.007	[−0.064, 0.051]	0.019	[−0.058, 0.095]	0.002	[−0.054, 0.058]
`timept[t18m]:STAI.T`	−0.033	[−0.11, 0.045]	0.11 *	[0.005, 0.21]	0.035	[−0.044, 0.11]
`emo[pos]:STAI.T`	−0.025	[−0.072, 0.022]				
`QSR.cert:STAI.T`	−0.009	[−0.026, 0.009]	−0.021	[−0.047, 0.005]	−0.002	[−0.027, 0.023]
`grp[PT]:timept[t18m]:emo[pos]`	0.027	[−0.16, 0.22]				
`grp[PT]:timept[t18m]:QSR.cert`	0.065	[−0.067, 0.20]	−0.14 †	[−0.31, 0.027]	0.031	[−0.11, 0.17]
`grp[PT]:emo[pos]:QSR.cert`	0.070	[−0.050, 0.19]				
`timept[t18m]:emo[pos]:QSR.cert`	0.081	[−0.027, 0.188]				
`grp[PT]:timept[t18m]:STAI.T`	0.029	[−0.067, 0.13]	−0.14 *	[−0.27, −0.008]	−0.045	[−0.14, 0.051]
`grp[PT]:emo[pos]:STAI.T`	0.038	[−0.035, 0.11]				
`timept[t18m]:emo[pos]:STAI.T`	0.12 *	[0.023, 0.22]				
`grp[PT]:QSR.cert:STAI.T`	0.003	[−0.028, 0.034]			−0.018	[−0.064, 0.028]
`timept[t18m]:QSR.cert:STAI.T`	−0.069 **	[−0.12, −0.022]			−0.059	[−0.13, 0.013]
`grp[PT]:timept[t18m]:emo[pos]:QSR.cert`	−0.16 *	[−0.32, −0.0004]				
`grp[PT]:timept[t18m]:emo[pos]:STAI.T`	−0.16 **	[−0.29, −0.041]				
`grp[PT]:timept[t18m]:QSR.cert:STAI.T`	0.053	[−0.011, 0.12]			0.092 †	[−0.006, 0.19]
Constant	0.050	[−0.041, 0.14]	−0.067	[−0.18, 0.051]	0.10 *	[0.012, 0.19]
marg,R^2 / cond.R^2	0.17 / 0.37		0.16 / 0.37		0.11 / 0.11	
N	160		80		80	
Log Likelihood	48.4		2.8		17.9	
AIC	−32.7		24.3		0.2	
BIC	65.7		60.1		43.1	

Predictors [reference level of predictor] separated by a colon (:) indicate interactions between those predictors. Predictors tested were group (`grp`: healthy control `HC`, breast cancer patient `PT`); testing time point (`timept`: before cancer treatments in PT (`t0m`), a year after chemotherapy or 18 months from `t0m` (`t18m`)); emotional valence of memory item (`emo`: positive `pos`, negative `neg`); QSR self-certitude subscore (`QSR.cert`); and trait anxiety score (`STAI.T`). Wald test $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$.

Table S3. Differences in volume and morphology (metric) across time points (Δ) between healthy controls (HC) and breast-cancer patients (PT). Changes between timepoints (dt) are shwon separately only for PT (right block of columns under ΔPT).

			ΔPT > ΔHC				ΔPT			
region	metric	side	Δspc	$t_{\sim 30-38}$	p_{FDR}	d	$\overline{\text{spc}}_{dt}$	$t_{dt,19}$	p_{FDR}	d
Hc (whole)	vol	L	0.025	0.046	–	–	−0.13	−0.46	0.98	–
		R	−0.59	−1.35	0.19	−0.43	−0.27	−0.85	0.85	–
aHc	**vol**	**L**	**−1.59**	**−2.21**	**0.033**	**0.70**	0.47	1.05	0.82	–
		R	0.18	−0.30	0.77	–	−0.96	−2.37	0.42	–
pHc	vol	L	0.79	1.29	0.20	0.41	−0.044	−0.11	0.99	–
		R	−0.76	−0.96	0.34	0.30	0.42	0.79	0.85	–
Amyg	vol	L	0.81	0.84	0.41	–	0.63	1.00	0.82	–
		R	0.46	0.36	0.72	–	−0.98	−1.50	0.80	–
Th	**vol**	**L**	**1.21**	**2.88**	**0.007**	**0.91**	**−1.32**	**−4.78**	**0.008**	**−1.07**
		R	0.071	0.14	0.89	–	0.14	0.43	0.98	–
ACC	*fold*	L	−0.55	−0.27	–	–	0.62	0.36	0.98	–
		R	*−2.89*	*−1.79*	*0.082*	*−0.57*	2.63	2.38	0.42	–
vmPFC	**vol.S**	**L**	**2.46**	**2.72**	**0.011**	**0.86**	−1.22	−1.58	0.80	–
		R	−1.16	−1.06	0.30	−0.34	0.43	0.56	0.98	–

All volumes (vol) and surface-based metrics such as folding index (fold) and surface-geometry-based volume (vol.S) were estimated by FreeSurfer's longitudinal volumetric parcellation approach except for anterior hippocampus (aHc) and posterior hippocampus (pHc) volumes which were estimated by ASHS. Two-sample 2-sided Welch's t-statistics with associated FDR-corrected p values and effect sizes in terms of Cohen's d (d) are shown as well as between testing point (dt) mean symmetrized percent change estimates ($\overline{\text{spc}}_{dt}$) for each group. A positive t-statistic for group comparisons indicates greater PT than HC change. A negative spc value indicates reduction in the given measure with time. Reliable differences are marked in bold.

S3.1.3 Negativity Effect (NE)

HC but not PT was found to show better memory for negative items judged with respect to a specific person (whether `SR` or `OR`) as opposed to merely judging the valence of a negative trait (`(SR+OR).neg>SEM.neg`, HC.t0m: $t_{39} = 2.5, p_{FDR} = 0.061, d = 0.40$; HC.t18m: $t_{39} = 2.9, p_{FDR} = 0.043, d = 0.46$; PT: all $|t_{39}| < 1, p_{FDR} > 0.50$). HC's better memory for negative items was due to the better memory for `SR.neg>SEM.neg` items only (HC.t0m: $t_{19} = 2.8, p_{FDR} = 0.046, d = 0.62$; HC.t18m: $t_{19} = 2.8, p_{FDR} = 0.042, d = 0.64$) while PT did not show an advantage for negative traits even when self-judged compared to general semantic judgment (all $|t_{19}| < 1.5, p_{FDR} > 0.31$).

For `SR.pos` items, PT showed better memory than for semantic judgment of positive traits (`SR.pos>SEM.pos`; PT.t0m: $t_{19} = 4.1, p_{FDR} = 0.006$, Cohen's $d = 0.91$; PT.t18m: $t_{19} = 2.3, p_{FDR} = 0.11$, Cohen's $d = 0.51$), while HC did show such a memory advantage for `SR.pos` but only at t18m (HC.t0m: $t_{19} = 1.1, p_{FDR} = 0.45$; HC.t18m: $t_{19} = 2.8, p_{FDR} = 0.043$, Cohen's $d = 0.63$).

S3.1.4 Reference and valence

Best fitting linear mixed models for the memory performance difference for self-judged positive over positive items judged of others (the difference $\text{gmean}_{\texttt{SR.pos}} - \text{gmean}_{\texttt{OR.pos}}$, or $\Delta\text{gmean}_{\texttt{SR.pos-OR.pos}}$) or for positive `SR` over negative `SR` items ($\Delta\text{gmean}_{\texttt{SR.pos-SR.neg}}$) also suggested that group, testing time point, QSR certitude score, STAI trait anxiety score and their interactions were weakly and moderately predictive, respectively (Table S2).

S3.1.5 Judgment response times

Subsequently remembered items were not because longer time was spent encoding/judging an item (Welch's 2-sample 2-sided t-tests; HC.t0m: $t_{442.0} = 0.59, p = 0.28$; HC.t18m: $t_{386.6} = 1.48, p = 0.070$; PT.t0m: $t_{368.9} = 0.72, p = 0.24$; PT.t18m: $t_{328.0} = -0.51, p = 0.69$). Compared to HC, PT were slower at judging all items, regardless of condition and emotional valence (Welch's 2-sample t-tests; all $t < -3.3, p < 0.001$). PT did not spend more time judging negatively over positively valenced character traits in any of the conditions at either time points (Welch's 2-sample 2-sided t-tests; largest statistic for `SEM.t0`: $t_{114.0} = 1.46, p_{FDR} = 0.15$; `SR.t0m`: $t_{121.7} = 0.58, p_{FDR} = 0.56$; `SR.t18m`: $t_{114.7} = -0.32, p_{FDR} = 0.75$). However, HC spent more time judging negative than positive traits in the `SEM` condition at t18m(est. 150 ms longer; $t_{113.7} = 4.23, p_{FDR} < 0.001, d = 0.78$), a little less so for `SR` questions also at t18m (est. 86 ms longer; $t_{119.4} = 2.45, p_{FDR} = 0.016, d = 0.44$).

S3.2 Structural brain changes

Table S3 shows the mean symmetrized percent brain structural changes for target regions and differences found between PT and HC, as well as over time in the PT group.

S3.3 Multivariate analysis

The clustering results which best separated PT and HC groups are shown in Fig. S3.

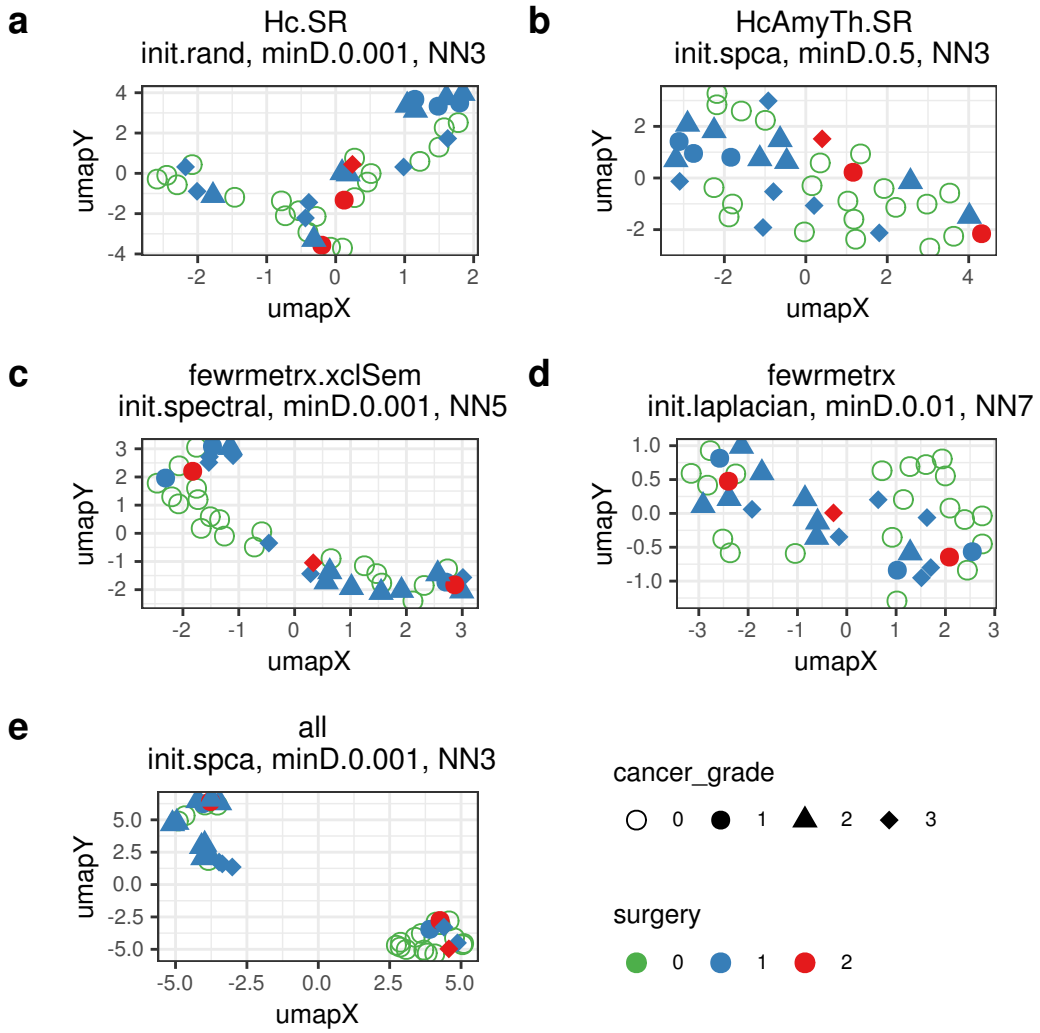


Figure S3. Best clustering results from exploratory analyses with UMAP on different feature/variable sets and algorithm parameters. All feature sets were symmetrized percent changes in structural and functional MRI metrics and anxiety and self-representation neuropsychological scores between testing time points. Feature sets used were nested, and plots are shown here in order of increasing feature set size. **a)** Shows the clusters found from the smallest feature set common to all other sets. It is comprised of volumetric changes, beta estimates from mass-univariate GLM of BOLD activation and connectivity (gPPI) analyses for the condition contrast `SR.H.pos>SR.H.neg` extracted from target ROIs (see Sec. S3.3 and main text), anxiety scores (STAI.S and STAI.T), self-representation scores (QSR.certitude and QSR.selfEsteem), and the completion time differences between TMT-A and TMT-B. **b)** Clusters found based on features in core set along with volumetrics for the amygdala and thalamic divisions. **c)** Clusters found for feature set including those in *b* along with volumetrics and surface-based metrics from all the regions from the standard atlases used (see main text), and beta estimates/gPPI estimates for `SR.H.pos>OR.H.pos`. **d)** Also includes `SR.H.[pos|neg]>SEM.H.[pos|neg]` contrasts. **e)** Clusters found for the maximal feature set which includes the remaining volumetrics and surface-based metrics provided by the chosen atlases (see main text). Data points (a participant) are shown by cancer grade and if participant had undergone surgery (0 = no surgery, 1 = lumpectomy, 2 = mastectomy. A cancer grade of 0 is no cancer (HC). As all PT underwent surgery, surgery = 0 is only HC.

S4 Additional Discussion: Limitations and Future Perspectives

Although the forte of this study was its longitudinal design, practical limitations prevented us from recruiting enough patients to focus on the different cognitive impact of surgical intervention, each of the adjuvant cancer treatments, as well as the impact of physiological changes during menopause. For example, hormonal therapy such as tamoxifen also has known impact on cognitive functioning including memory[35]. In our study, all our PT were undergoing hormonal therapy by t18m. To be able to generalize our findings and answer our followup questions of the factors that compromise episodic memory abilities and discrepancies of these abilities across test measures and between tested measures and patient complaints, a larger sample of PT with a wider range of anxiety and self-certitude would be required. Healthy participants should not only be matched by age as well as anxiety and self-certitude. Our findings also raise the point that episodic memory and events in day-to-day life are more often emotionally laden, especially with breast cancer patients whose situation can often involve a re-evaluation of themselves and their lives. Even in our limited sample of PT who had no clinically relevant anxiety or depression, there are indications that memory resources are regulated differently compared to our HC. Studies in memory in cancer patients often focus on classic memory paradigms testing working memory, such as n-back working memory task. Such tasks have limited generalizability to relevant emotional re-evaluations[36].

This leads us to offering recommendations on task design in future studies to better understand patient's cognitive complaints and the heterogeneity of their problems. Episodic memory paradigms could benefit from being able to dissociate individually reported arousal from emotional valence. Emotional regulation difficulties are related to attention-related problems and perceived stress[37]. The Hc, known to be susceptible to elevated cortisol levels[38,39], whether due to inflammatory responses or psychological stress, can suffer more damage, therefore contributing to the multifactorial problem contributing to the elusiveness of the extent of the cognitive problems patients report. Neutral-valenced and non-arousing items and their permutations would be advantageous to measure attentional disengagement due to arousal. For example, as we did not have any neutral emotional valence items, we have no comparison to conclude whether slower RTs could also be due to emotional avoidance. The task must also be feasible but still challenging enough to generate enough incorrect memory responses to contrast remembered from forgotten items. The most power-efficient approach would be to individually adjust task difficulty during pre-test sessions.

References

1. Duval, C., Eustache, F. & Piolino, P. Multidimensional self, autobiographical memory and aging. *Psychol Neuropsychiatr Vieil* **5**, 179–92 (2007).
2. Duval, C. *et al.* What happens to personal identity when semantic knowledge degrades? a study of the self and autobiographical memory in semantic dementia. *Neuropsychologia* **50**, 254–265, DOI: 10.1016/j.neuropsychologia.2011.11.019 (2012).
3. Spielberger, C. D. State-trait anxiety inventory for adults (STAI-AD). Tech. Rep., APA PsycTests (1983). DOI: 10.1037/t06496-000.
4. Chételat, G. *et al.* Dissociating atrophy and hypometabolism impact on episodic memory in mild cognitive impairment. *Brain* **126**, 1955–1967, DOI: 10.1093/brain/awg196 (2003).
5. Eustache, F., Desgranges, B. & Lalevée, C. [clinical evaluation of memory]. *Rev Neurol (Paris)* **154**, S18–32 (1998). PMID: 9834542.
6. Fouquet, M. *et al.* Role of hippocampal CA1 atrophy in memory encoding deficits in amnestic mild cognitive impairment. *NeuroImage* **59**, 3309–3315, DOI: 10.1016/j.neuroimage.2011.11.036 (2012).
7. Corrigan, J. D. & Hinkeldey, N. S. Relationships between parts A and B of the Trail Making Test. *J Clin Psychol* **43**, 402–409, DOI: 10.1002/1097-4679(198707)43:4<402::AID-JCLP2270430411>3.0.CO;2-E (1987).
8. Cohen, J. *Statistical power analysis for the behavioral sciences* (Lawrence Erlbaum Associates, New York, 1988), second edn.
9. Hainmueller, J., Mummolo, J. & Xu, Y. How much should we trust estimates from multiplicative interaction models? simple tools to improve empirical practice. *Polit. Analysis* **27**, 163—192, DOI: 10.1017/pan.2018.46 (2019).
10. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat Methods* DOI: 10.1038/s41592-018-0235-4 (2018).
11. Esteban, O. *et al.* fmriprep. *Software* DOI: 10.5281/zenodo.852659 (2018).
12. Gorgolewski, K. *et al.* Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front Neuroinform* **5**, 13, DOI: 10.3389/fninf.2011.00013 (2011).
13. Gorgolewski, K. J. *et al.* Nipype. *Software* DOI: 10.5281/zenodo.596855 (2018).
14. Tustison, N. J. *et al.* N4ITK: Improved N3 bias correction. *IEEE Trans Med Imaging* **29**, 1310–1320, DOI: 10.1109/TMI.2010.2046908 (2010).
15. Avants, B., Epstein, C., Grossman, M. & Gee, J. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Analysis* **12**, 26–41, DOI: 10.1016/j.media.2007.06.004 (2008).
16. Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* **20**, 45–57, DOI: 10.1109/42.906424 (2001).
17. Reuter, M., Rosas, H. D. & Fischl, B. Highly accurate inverse consistent registration: A robust approach. *NeuroImage* **53**, 1181–1196, DOI: 10.1016/j.neuroimage.2010.07.020 (2010).
18. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis: I. segmentation and surface reconstruction. *NeuroImage* **9**, 179–194, DOI: 10.1006/nimg.1998.0395 (1999).
19. Klein, A. *et al.* Mindboggling morphometry of human brains. *PLoS Comput. Biol* **13**, e1005350, DOI: 10.1371/journal.pcbi.1005350 (2017).
20. Fonov, V., Evans, A., McKinstry, R., Almli, C. & Collins, D. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage* **47, Supplement 1**, S102, DOI: 10.1016/S1053-8119(09)70884-5 (2009).
21. Evans, A., Janke, A., Collins, D. & Baillet, S. Brain templates and atlases. *NeuroImage* **62**, 911–922, DOI: 10.1016/j.neuroimage.2012.01.024 (2012).
22. Greve, D. N. & Fischl, B. Accurate and robust brain image alignment using boundary-based registration. *NeuroImage* **48**, 63–72, DOI: 10.1016/j.neuroimage.2009.06.060 (2009).
23. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* **17**, 825–841, DOI: 10.1006/nimg.2002.1132 (2002).

24. Cox, R. W. & Hyde, J. S. Software tools for analysis and visualization of fmri data. *NMR Biomed* **10**, 171–178, DOI: 10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L (1997).

25. Pruim, R. H. R. *et al.* ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *NeuroImage* **112**, 267–277, DOI: 10.1016/j.neuroimage.2015.02.064 (2015).

26. Power, J. D. *et al.* Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage* **84**, 320–341, DOI: 10.1016/j.neuroimage.2013.08.048 (2014).

27. Behzadi, Y., Restom, K., Liau, J. & Liu, T. T. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage* **37**, 90–101, DOI: 10.1016/j.neuroimage.2007.04.042 (2007).

28. heodore D. Satterthwaite *et al.* An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage* **64**, 240–256, DOI: 10.1016/j.neuroimage.2012.08.052 (2013).

29. Lanczos, C. Evaluation of noisy data. *SIAM J Numer Anal* **1**, 76–85, DOI: 10.1137/0701007 (1964).

30. Abraham, A. *et al.* Machine learning for neuroimaging with scikit-learn. *Front Neuroinform* **8**, DOI: 10.3389/fninf.2014.00014 (2014).

31. Destrieux, C., Fischl, B., Dale, A. & Halgren, E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage* **53**, 1–15, DOI: 10.1016/j.neuroimage.2010.06.010 (2010).

32. Fischl, B. *et al.* Automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341–355, DOI: 10.1016/S0896-6273(02)00569-X (2002).

33. Esteban, O. *et al.* MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS ONE* **12**, 1–21, DOI: 10.1371/journal.pone.0184661 (2017).

34. Vos, P. J., Garssen, B., Visser, A. P., Duivenvoorden, H. J. & de Haes, H. C. J. M. Early stage breast cancer: Explaining level of psychosocial adjustment using structural equation modeling. *J Behav Med* **27**, 557–580, DOI: 10.1007/s10865-004-0003-z (2004).

35. Chen, X. *et al.* Decision-making impairments in breast cancer patients treated with tamoxifen. *Horm Behav* **66**, 449–456, DOI: 10.1016/j.yhbeh.2014.07.005 (2014).

36. Lee, T.-W. & Xue, S.-W. Does emotion regulation engage the same neural circuit as working memory? a meta-analytical comparison between cognitive reappraisal of negative emotion and 2-back working memory task. *PLoS ONE* **13**, 1–25, DOI: 10.1371/journal.pone.0203753 (2018).

37. Shahane, A. D., Lopez, R. B. & Denny, B. T. Implicit reappraisal as an emotional buffer: Reappraisal-related neural activity moderates the relationship between inattention and perceived stress during exposure to negative stimuli. *Cogn Affect. Behav Neurosci* **19**, 355–365, DOI: 10.3758/s13415-018-00676-x (2019).

38. Lucassen, P. J. *et al.* Neuropathology of stress. *Acta Neuropathol* **127**, 109–135, DOI: 10.1007/s00401-013-1223-5 (2014).

39. Frodl, T. & O'Keane, V. How does the brain deal with cumulative stress? a review with focus on developmental stress, hpa axis function and hippocampal structure in humans. *Neurobiol Dis* **52**, 24–37, DOI: 10.1016/j.nbd.2012.03.012 (2013).