

Research Articles: Behavioral/Cognitive

## Ventromedial prefrontal cortex drives the prioritization of self-associated stimuli in working memory

https://doi.org/10.1523/JNEUROSCI.1783-20.2020

Cite as: J. Neurosci 2021; 10.1523/JNEUROSCI.1783-20.2020

Received: 9 July 2020 Revised: 14 October 2020 Accepted: 15 November 2020

This Early Release article has been peer-reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.jneurosci.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Copyright © 2021 the authors

1	Ventromedial prefrontal cortex drives the prioritization of
2	self-associated stimuli in working memory
3	Abbreviated tittle: VMPFC drives self-prioritization in working memory
4	
5	Shouhang Yin <sup>1,2</sup> , Taiyong Bi <sup>3</sup> , Antao Chen <sup>2*</sup> , Tobias Egner <sup>4</sup>
6	
7	<sup>1</sup> School of Mathematics and Statistics, Southwest University, Chongqing 400715,
8	China.
9	<sup>2</sup> Key laboratory of Cognition and Personality of the Ministry of Education, Faculty of
10	Psychology, Southwest University, Chongqing 400715, China
11	<sup>3</sup> Center for Mental Health Research in School of Management, Zunyi Medical
12	University, Guizhou 563006, China
13	<sup>4</sup> Center for Cognitive Neuroscience, and Department of Psychology and Neuroscience,
14	Duke University, Durham, NC 27708, USA
15	<sup>*</sup> Correspondence should be addressed to: Dr. Antao Chen, Faculty of Psychology,
16	Southwest University, Chongqing City, China. E-mail: xscat@swu.edu.cn
17	

- <u>JNeurosci Accepted Manuscript</u>
- 19 The paper includes 53 pages, 7 figures, and 4 tables.
- 20 Word counts: abstract = 172, introduction = 646, discussion = 1499
- 21
- 22 Acknowledgements: This work was supported by grants from the National Natural
- 23 Science Foundation of China (31771254), and China Postdoctoral Science Foundation
- 24 (2019M663425), and the fellowship of China National Postdoctoral Program for
- 25 Innovative Talents (BX20200283).
- 26

Author contributions: S. Yin and A. Chen initially conceived and designed the
experiments, which were later improved and modified by discussions among A. Chen
and T. Egner. S. Yin and T. Bi performed the experiments. The data were analyzed by
S. Yin, T. Bi, and A. Chen, and the paper was written by S. Yin, T. Bi, and A. Chen,
and T. Egner.

32

33 Competing interests: The authors declare no competing financial interests.

## 35 Abstract

36	Humans show a pervasive bias for processing self- over other-related information,
37	including in working memory (WM), where people prioritize the maintenance of self-
38	(over other-) associated cues. To elucidate the neural mechanisms underlying this
39	self-bias, we paired a self- vs. other-associated spatial WM task with functional
40	magnetic resonance imaging (fMRI) and transcranial direct current stimulation (tDCS)
41	of human participants of both sexes. Maintaining self- (over other-) associated cues
42	resulted in enhanced activity in classic WM regions (frontoparietal cortex), and in
43	superior multivoxel pattern decoding of the cue locations from visual cortex.
44	Moreover, ventromedial prefrontal cortex (VMPFC) displayed enhanced functional
45	connectivity with WM regions during maintenance of self-associated cues, which
46	predicted individuals' behavioral self-prioritization effects. In a follow-up tDCS
47	experiment, we targeted VMPFC with either excitatory (anodal), inhibitory (cathodal),
48	or sham tDCS. Cathodal tDCS eliminated the self-prioritization effect. These findings
49	provide strong converging evidence for a causal role of VMPFC in driving
50	self-prioritization effects in WM and provide a unique window into the interaction
51	between social, self-referential processing and high-level cognitive control processes.
52	
53	Keywords: self-reference, self-prioritization, ventral medial prefrontal cortex,
54	working memory
55	

## 56 Significance statement

57	People have a strong tendency to attend to self-related stimuli, such as their names.
58	This self-bias extends to the automatic prioritization of arbitrarily self-associated
59	stimuli held in working memory. Since working memory is central to high-level
60	cognition, this bias could influence how we make decisions. It is therefore important
61	to understand the underlying brain mechanisms. Here, we used neuroimaging and
62	noninvasive neurostimulation techniques to show that the source of self-bias in
63	working memory is the ventromedial prefrontal cortex, which modulates activity in
64	frontoparietal brain regions to produce prioritized representations of self-associated
65	stimuli in sensory cortex. This work thus reveals a brain circuit underlying the
66	socially motivated (self-referential) biasing of high-level cognitive processing.

## 67 Introduction

68	People show a pervasive bias toward preferentially processing self-related
69	information compared with other-related information. For instance, intrinsically
70	self-related stimuli, like one's name or face, are prioritized in long-term memory
71	(Kesebir and Oishi, 2010), attract attention more potently (Alexopoulos et al., 2012;
72	Liu et al., 2016), and are perceived quicker and more faithfully than other-related
73	stimuli (Sui et al., 2012). We have recently shown that this type of self-prioritization
74	is evident even in working memory (WM) (Yin et al., 2019), the mental workspace
75	where information is temporarily maintained and manipulated to guide behavior
76	(D'Esposito and Postle, 2015). When people had to keep in mind different spatial
77	locations, they prioritized the WM maintenance of those locations where (arbitrary)
78	self-associated cues compared to other-associated cues had been presented, even
79	though self-associated stimuli were no more likely to be probed than other-associated
80	stimuli (Yin et al., 2019).
81	Understanding the processes underlying this form of social biasing of high-level
82	cognition has important implications, as WM representations are central to
83	decision-making and cognitive control (Gazzaley and Nobre, 2012; Boureau et al.,
84	2015). To gain a deeper understanding of how WM representations are biased towards
85	self-associated information, we paired a self- vs. other-associated spatial WM task
86	(Yin et al., 2019) with functional magnetic resonance imaging (fMRI) and transcranial

87	direct current stimulation (tDCS). Specifically, participants were first trained to form
88	associations between three colors and three persons: one with themselves, one with a
89	best friend, and the third with a stranger. Then, they performed a delayed
90	match-to-sample spatial WM task where they needed to memorize the locations and
91	social labels of two color cues and then completed a recognition test. We tested two
92	key neural predictions, derived from the literature (see below): (1) the behavioral
93	effect of self-prioritization in WM would be mirrored by enhanced activity for, and
94	more faithful representation of, self-associated items in brain regions supporting WM;
95	and (2), this effect would arise from the influence on WM regions by brain areas
96	specialized for processing self-related information.
97	First, a large neuroimaging literature has outlined a WM network consisting of
98	dorsolateral prefrontal cortex, the frontal eye field [FEF]), and posterior parietal
99	cortex (PPC), including the intraparietal sulcus (IPS) and superior parietal lobule
100	(SPL) (Baluch and Itti, 2011; Petersen and Posner, 2012). If self-associated stimuli
101	were afforded special priority in WM, we would expect activity in these regions to be
102	enhanced when keeping self- compared to other-associated items in mind. Moreover,
103	the currently predominant sensory recruitment hypothesis of WM (Serences, 2016;
104	Scimeca et al., 2018) holds that frontoparietal cortex is responsible for activating (or
105	attending to) representations of WM items, but that those representations are
106	maintained in - and thus decodable from - sensory cortex (Sprague et al., 2014;
107	Rahmati et al., 2017; Cai et al., 2019; Rademaker et al., 2019). Accordingly, we

108	expected the decoding of WM cue locations from activity patterns in visual cortex to
109	be superior for self- than for other-associated cue locations.
110	Second, previous studies have consistently implicated midline structures of the
111	ventral medial prefrontal cortex (VMPFC) and the posterior cingulate cortex (PCC),
112	key nodes of the default mode (DM) network(Raichle, 2015), when contrasting self-
113	with other-referential processing (Qin et al., 2012; Sui et al., 2013; Yankouskaya et al.,
114	2017). We expected to replicate this finding here in the domain of WM. Moreover, we
115	expected that these self-referential processing regions would exhibit increased
116	functional coupling with WM-related regions during the maintenance of self-
117	compared to other-associated items, reflecting the hypothesized biasing of the WM
118	network. Finally, based on fMRI results conforming to the above predictions, we
119	performed a follow-up tDCS experiment where we targeted VMPFC in three
120	independent groups of participants who received anodal, cathodal, or sham
121	stimulation. If VMPFC contributed causally to the self-prioritization effect in WM,
122	we would expect to see this effect enhanced under anodal compared to sham
123	stimulation or diminished under cathodal compared to sham stimulation.
124	Materials and Methods

Participants. Thirty-four participants took part in the fMRI study. Out of those 34,
two terminated the scan prematurely, and data from four other participants were
excluded due to excessive head motion (three participants, more than 3 mm or 3

128	degrees) or poor WM task performance (one participant, lower than 80%). Another
129	two participants were excluded only from the visual cortex decoding analysis, due to
130	excessive head motion during the retinotopic mapping scan (more than 3 mm or 3
131	degrees). Thus, after exclusion, 28 participants (11 females, mean age = 20.47 years,
132	S.D = 0.97 years) remained for the main fMRI data analyses, and 26 participants (10
133	females, mean age = $20.50$ years, S.D = $1.00$ years) remained for the visual cortex
134	decoding analysis. Ninety new participants were recruited for the tDCS study, and
135	split into three groups: anode (15 females, mean age = $20.85$ years, S.D = $1.45$ years),
136	cathode (15 females, mean age = $21.18$ years, S.D = $1.61$ years), and sham (15
137	females, mean age = $20.89$ years, S.D = $1.74$ years). All participants were
138	right-handed with reported normal or corrected-to-normal vision and had no known
139	neurological or visual disorders. Both experiments were approved by the University
140	Human Ethics Committee of Southwest University (China). All volunteers gave
141	informed written consent and were compensated for their participation.
142	Stimuli and procedure of fMRI WM task. The full timeline of the procedure of the
143	present study is presented in Figure 1a. Before entering the scanner, participants
144	partook in an associative learning procedure (Sui et al., 2012; Yin et al., 2019).
145	Participants were initially instructed to associate one of the colors with the self, one
146	with a named best friend, and one with an unfamiliar person for 60 seconds. These
147	associations were counterbalanced across participants and subsequently employed in

148	the spatial WM task in the scanner. This approach of creating novel color-self/other
149	associations avoids the confounding impact of familiarity on self-reference effects
150	(Sui et al., 2012) and thus allowed us to probe self-prioritization in WM in a tightly
151	controlled manner. Then, participants performed a color-label matching task, where
152	on each trial a circle $(1.2^{\circ} \times 1.2^{\circ})$ in one of the three colors was presented above a
153	black fixation cross at the center of a gray screen. One of three possible Chinese
154	characters (for self, friend, or stranger, $2.4^{\circ}/3.4^{\circ} \times 1.2^{\circ}$ ) was displayed below the
155	fixation cross. The visual angle between the center of the colored circle or the word
156	and the fixation cross was 3.5°. Participants had to indicate whether the color-label
157	pairing matched with the instructed association, using the index and middle fingers of
158	the right hand on the keypad keys "1" and "2". Each trial started with a 500 ms
159	fixation cross, followed by a 200 ms pairing probe, after which a blank screen was
160	presented and participants had 1500 ms to press a key as quickly and accurately as
161	possible. The presentation of the blank screen was terminated by key press or after
162	1500 ms, and the trial ended with a 500 ms feedback display. Each participant
163	performed a block of 30 trials, and their accuracy had to be at least 80% to move on to
164	the next phase of the study. The matching task served as training to make participants
165	master the color-label associations.
166	

**Figure 1** 



169 Figure 1. Task protocol and example stimuli, a The overall experiment procedure consisted of a 170 learning phase, a training phase, the fMRI WM task, and a subsequent retinotopic mapping scan. b 171 Example stimuli and timing of presentation of a single trial of the WM task. Participants had to 172 remember the locations of two different color cues (previously associated with different social labels), 173 each of which could occur in one of four locations (one cue per visual hemifield), as indicated by the 174 dotted circle placeholders. After an 8 s delay, they responded (yes/no) to a WM probe shown at one of 175 the locations. If the trial was a match trial, the location probe response was followed by a verbal probe 176 for the social label associated with the color (e.g., "friend"), to which the participant had to give 177 another yes/no response. The unit of the numbers under horizontal axis is second.

179	The fMRI task was a delayed match-to-sample spatial WM task adapted from
180	our previous behavioral study (Yin et al., 2019). As displayed in Figure 1b, on a gray
181	background, on each trial two different-colored cues (filled-in circles, subtending 1.2°
182	$\times$ 1.2° of visual angle) were presented, one to the left and one to the right of a central
183	fixation cross, in one of four possible locations. Figures 1 and 3 show the eight
184	possible locations (four at each side of the visual field; bilateral symmetry). Two
185	possible cue locations were located horizontally parallel with the fixation cross, with
186	distances from fixation of $3.4^{\circ}$ and $4.6^{\circ}$ , respectively; the other two possible cue
187	locations were above and below the cue that was horizontally in line with, and the
188	closest to, the fixation cross, with vertical distances of 1.2°. A trial started with a

189	1000-ms fixation cross that remained on screen throughout the trial, followed by two
190	colored, filled-in circles shown for 1000 ms. Participants were asked to remember the
191	locations and social labels associated with these color cues (based on the prior
192	learning task). Then the trial entered an 8000 ms delay-period, after which the font of
193	the fixation-cross turned bold for 300 ms (signaling the end of the delay period). A
194	WM probe (a black filled-in circle) was then presented for 1500 ms at one of the eight
195	possible locations, and the participants had to judge whether the probe location
196	matched either of the two remembered cue locations, using the index and middle
197	fingers of their right hand to indicate yes or no. The assignment of response finger to
198	responses was counterbalanced across participants.
199	The WM probe presentation was terminated by the key press or after 1500 ms,
200	after which an adjustable duration blank screen interval was presented to keep the
201	entire target plus blank screen presentation time at 2000 ms. If the probe matched
202	either of the two remembered locations (match trial) and the participant indicated this
203	correctly, a label word (Self, Friend, Stranger) was presented at the probe location for
204	1500 ms, and participants were required to judge whether the label word matched the
205	remembered color in this location. Probing the color-label after match trials served to
206	ensure that participants kept actively remembering the social labels associated with
207	each color (not just the colors). Following the response, another adjustable duration

209	screen at 2000 ms. On non-match trials, only a 2000-ms blank screen was presented.
210	Finally, each trial ended with a (baseline) blank screen presentation of 4000 ms.
211	The different possible combinations of the color memory cues resulted in three
212	trial types or pairings: Self-Friend, Self-Stranger, and Friend-Stranger. For instance, a
213	Self-Friend trial may present the self-associated color cue in one of the left-hand
214	locations and the friend-associated color cue in one of the right-hand locations. Each
215	of these trial types occurred 64 times, including 16 match trials for each of the two
216	items and 32 non-match trials. Altogether, there were 192 trials, including 32
217	self-match trials, 32 friend-match trials, 32 stranger-match trials and 96 non-match
218	trials, evenly broken down into 8 runs (each trial type or matching type occurred
219	equally often in each run); all kinds of trials were presented in pseudorandom order.
220	Our study was specifically designed to facilitate decoding of the (self- or
221	other-associated) cue locations from fMRI data in visual cortex, by always presenting
222	one cue per visual hemifield, and by acquiring a retinotopic mapping and WM cue
223	location localizer scan: A standard phase-encoded method developed by Sereno et al.
224	was used to define retinotopic visual areas (Sereno et al., 1995), in which participants
225	viewed a rotating wedge that created traveling waves of neural activity in visual
226	cortex (2 runs). Another independent block-design localizer run was performed to
227	localize the retinotopic area where the stimuli were presented in the WM task. In this
228	run, to localize regions in visual cortex responsive to the visual field locations where

the targets could appear, two flickering triangular checkerboards covering the edges of
possible stimuli locations were presented on each side of the screen for 12 seconds.
The run contained 14 checkerboard blocks, interleaved with blank screen blocks of 12
seconds.

233 Experimental Design and Statistical Analysis. As detailed above, there were three 234 possible combinations of the color memory cues: Self-Friend, Self-Stranger, and 235 Friend-Stranger, and there were three types of location probe match response: 236 self-match, friend-match, and stranger-match. In the behavioral analysis, our focus 237 was the response times (RTs) of the location probe match trials. Thus, the task is a 238 3-level single-factor within-subjects design and a repeated-measures one-way analysis 239 of variance (ANOVA) was performed on the RT data. In univariate neuroimaging 240 analyses, two effects were examined, one using a contrast to identify self-associated 241 activation (self contrast: Self-Friend > Friend-Stranger conditions), and the other one 242 to delineate regions involved in WM maintenance (WM contrast: contrasting 243 delay-period activity for Self-Friend, Self-Stranger, and Friend-Stranger trials > 244 baseline); both effects were analyzed with t-tests, using correction for multiple 245 comparisons. In the multivoxel pattern analysis (MVPA), trials were divided into two 246 groups: one where the self-associated cue was presented in the left visual hemifield 247 and the other-associated cue in the right visual hemifield (Self L trials), and the other 248 one corresponding to the opposite scenario (Self R trials). For each group of trials,

249	MVPAs were conducted on every time point of a trial to decode the four possible WM
250	cue locations, and the decoding accuracies were compared between self- and
251	other-associated cues using t-tests, corrected for multiple comparisons. In the
252	psychophysiological interaction (PPI) analysis, the VMPFC area activated in the
253	univariate self contrast was saved as a seed region mask, and the WM regions
254	activated in the univariate WM univariate contrast were saved as a target region mask.
255	The vector of the psychological variable of interest (Self-Friend > Friend-Stranger)
256	was calculated to create the psychophysiological interaction term, and neural
257	correlates of that interaction term were identified via a t-test, corrected for multiple
258	comparisons. We also used dynamic causal modeling (DCM) analysis to evaluate the
259	direction of influences between VMPFC and WM regions. Rival models were
260	evaluated statistically via Bayesian model comparison. The behavioral task in tDCS
261	experiment was identical to the fMRI WM task expect a reduction of the duration of
262	delay period, and the tDCS experiment is a 3 (Group: excitatory, inhibitory, and sham;
263	between-subjects) $\times$ 3 (Self-reference: self-match, friend-match, and stranger-match;
264	within-subjects) mixed design. All the statistical analyses were performed with SPSS
265	version 22.0 (SPSS, Inc., Chicago, IL, USA). Finally, summary behavioral and
266	neuroimaging data from this study can be accessed at
267	https://osf.io/jdwcr/?view_only=efdea02d46b1499d9c8db8692b175279.

268	<b>fMRI acquisition.</b> The WM task was run on a PC with an 18.5-in. monitor (1,366 $\times$
269	768 at 60 Hz), using E-prime software (Version 2.0), and participants watched the
270	screen through a mirror in the magnetic bore. Images were acquired with a Siemens
271	3T scanner (Siemens Magnetom Trio TIM, Erlangen, Germany), using a standard
272	12-channel radio-frequency head coil. An echo-planar imaging (EPI) sequence was
273	used for the collection of functional WM task data, and 221 T2-weighted images were
274	recorded per run (TR: 2000 ms; TE: 30 ms; flip angle: $85^{\circ}$ ; FoV: 224 × 224 mm <sup>2</sup> ;
275	matrix size: $64 \times 64$ ; in-plane resolution: $3.5 \times 3.5 \text{ mm}^2$ ; slice skip: 0.3 mm; 32
276	ascending 3 mm-thick slices). The retinotopic visual mapping and stimulus location
277	localizer scans were performed on the next day after the WM task scan, and signals
278	were acquired with an EPI sequence (TR: 2000 ms; TE: 30 ms; flip angle: 90°; FoV:
279	$192 \times 192 \text{ mm}^2$ ; matrix size: $64 \times 64$ ; in-plane resolution: $3.0 \times 3.0 \text{ mm}^2$ ; $33$
280	interleaved 3 mm-thick slices; no slice skip). The bottom slice was positioned at the
281	bottom of the temporal lobe. A high-resolution 3D structural data set (3D MPRAGE;
282	TR: 2600 ms; TE: 3.02 ms; flip angle: 8°; resolution: $1 \times 1 \times 1$ mm <sup>3</sup> ; 176 slices) was
283	collected before the retinotopic visual mapping scan.
284	fMRI data pre-processing. Image preprocessing and analysis were conducted in
285	Statistical Parametric Mapping toolbox (SPM12, Welcome Department of Imaging
286	Neuroscience, Institute of Neurology, London). The first five images were discarded

287 to achieve magnet-steady images. The imaging data were spatially realigned, and six

288	head motion parameters were estimated for inclusion in the task models. Images were
289	temporally realigned to the middle slice to correct for differences in slice timing.
290	Head motion within any MRI session was less than 3 mm or 3 degrees for any subject.
291	To normalize the functional images, each subject's structural brain image was
292	coregistered to the mean functional image and was subsequently segmented. The
293	parameters obtained in segmentation were used to normalize each subject's functional
294	image onto the Montreal Neurological Institute space (resampling voxel size: 3 mm <sup>3</sup> ).
295	A filter of 8 mm FWHM (full-width at half maximum) was used to spatially smooth
296	the normalized data.
297	General Linear Model (GLM) for fMRI data. A GLM approach was used to
298	estimate parameter values for event-related responses. Onsets of the retention period
299	were extracted for three trial types and the time series data were modeled for three
300	different vectors, corresponding to Self-Friend, Self-Stranger, and Friend-Stranger
301	conditions, respectively. Three additional regressors also modeled the respective
302	probe epochs for these conditions to control for their influence on retention period
303	activation estimates; another regressor modeled the blank screen stage as a no-task
304	baseline. The design matrices also included six head movement parameters to account
305	for any residual movement-related effect. All these vectors were convolved with the
306	canonical HRF. A high-pass filter was implemented with a cut-off of 128 seconds to
307	remove low-frequency drift from the time-series.

308	For each subject, we defined self contrast between Self-Friend and Friend-Stranger to
309	examine brain activation in relation to the self-prioritization effect, and another WM
310	contrast between the three conditions and the blank screen baseline to characterize
311	generic WM brain activation. These contrasts were then subjected to group-level
312	one-sample <i>t</i> -tests where participants were treated as random effects. Group analyses
313	were carried out within a grey matter mask to reduce total search space. For the
314	self-prioritization effect, we used a false discovery rate (FDR) to correct for multiple
315	comparisons in self contrast, with a voxelwise FDR-corrected threshold of $p < 0.05$
316	and an extent threshold of 30 voxels. This correction approach, which is more liberal
317	than a family-wise error correction, was chosen in order to gain greater sensitivity for
318	detecting potential effects in regions associated with self-referential processing that -
319	as part of the DM network – would normally be expected to be relatively suppressed
320	during a WM task. As the contrast of WM activity > baseline resulted in very broadly
321	distributed activity, and we were interested in only the most activated (core WM
322	network) regions, we subjected it to a more conservative correction method, with a
323	voxelwise FDR-corrected threshold of $p < 0.001$ and an extent threshold of 50 voxels.
324	To identify overlapping regions, we also performed a conjunction analysis by
325	overlapping the two contrast maps resulting from the above analyses. To examine the
326	activation patterns in regions showing both WM and self-prioritization effects in more
327	detail, we extracted the beta-values from these regions for each condition, using the
328	MarsBaR toolbox in SPM12.

329	Multivariate analysis for fMRI data. MVPAs were conducted using PRoNTo, a
330	pattern recognition toolbox for neuroimaging (http://www.mlnl.cs.ucl.ac.uk/pronto)
331	(Schrouff et al., 2013). Our primary MVPA was concerned with decoding the WM cue
332	locations from visual regions of interest based on the retinotopic mapping and WM
333	location localizer data. The anatomical volume for each subject was transformed into
334	the anterior commissure-posterior commissure (AC-PC) space (Talairach space).
335	Functional volumes of retinotopic mapping scans were preprocessed using
336	BrainVoyager QX, including 3D motion correction, linear trend removal, and
337	high-pass filtering (0.015 Hz). Head motion within any MRI session was less than 3
338	mm or 3 degrees for any subject. The functional volumes were then aligned to the
339	anatomical volume and transformed into the AC-PC space. Next, voxels were selected
340	for the MVPA based on their maximal responsiveness to both the retinotopic mapping
341	visual field localizer and the WM stimulus localizer task (see Stimuli and Procedure
342	for details). The 120 voxels (60 for each hemispheres) in primary visual cortex (V1)
343	that displayed the highest responses (gauged via t-statistics) to both localizers were
344	selected, and preprocessed but unsmoothed data were used for classifier training. The
345	left V1 voxels were trained to decode the locations of items that appeared on the right
346	field of vision, and vice versa for the right V1. This decoding analysis was conducted
347	on trials that contained self-associated WM cues, thus only including Self-Friend
348	trials and Self-Stranger trials, but no Friend-Stranger trials. These trials were divided
349	into two groups: one where the self-associated cue was presented in the left visual

350	hemifield and the other-associated cue in the right visual hemifield (Self_L, 64 trials);
351	and the other where the self-associated cue was presented in the right and the
352	other-associated cue is in the left hemifield (Self_R, 64 trials). There were four
353	possible cue locations on each side, and each location displayed 16 times in Self_L or
354	Self_R trials. Four classification analyses were conducted: left V1 for self-associated
355	cues, left V1 for other-associated cues, right V1 for self-associated cues, and right V1
356	for other-associated cues. In the present task, each trial contained 9 time points (TRs);
357	accordingly, the data of each time point were used as samples once, and four
358	classifications were conducted nine times, one per time point. All decoding analyses
359	were performed on single-subject data, with statistical reliability subsequently
360	assessed across the sample. Classification was accomplished using a multiclass
361	gaussian process, and classifier sensitivity was examined using a
362	leave-one-trial-per-group-out approach. Specifically, the classification prediction was
363	performed 16 times, and 60 trials (15 trials for each location) were used as training
364	data, leaving one trial for each location as the test trials. The significance of classifier
365	performance was determined using two-tailed, one-sample t-tests, testing against
366	chance performance of 0.25 ( $p < 0.05$ after FWE correction).
367	PPI and DCM analysis for fMRI data. PPI analyses were conducted using SPM12.
368	Based on the results of GLM, the VMPFC area activated in the Self-Friend >

369 Friend-Stranger contrast was saved as a seed region mask, and the (mostly

370	frontoparietal) regions activated in the WM contrast were saved as a target region
371	mask. For each subject, the exact VMPFC seed coordinate was defined using the peak
372	voxel in the individual first-level contrast between Self-Friend and Friend-Stranger
373	within the group mask. A sphere with a 6 mm radius was positioned at that peak of
374	each subject, and the deconvolved time course of VMPFC activity in this ROI was
375	extracted to serve as the physiological variable of interest.
376	The vector of the psychological variable of interest (Self-Friend >
377	Friend-Stranger) was calculated to create the psychophysiological interaction term
378	(the cross-product of the physiological and psychological variables). New SPMs were
379	computed for each subject, including the interaction term, the physiological variable
380	(that is, the VMPFC activation time course) and the psychological variable, as well as
381	six head movement parameters. We then identified brain regions within the WM-mask
382	where activation was predicted by the psychophysiological interaction term, reflecting
383	a change in functional coupling with the VMPFC as a function of condition (self- vs.
384	other-associated). The VMPFC activity and the psychological regressors were treated
385	as confound variables. Afterwards, individuals' contrast images were entered into a
386	group one-sample t-test where participants were treated as random effects, and
387	assessed for significance using a FDR-corrected threshold of $p < 0.05$ .
388	PPI analysis cannot provide evidence concerning the direction of functional
389	interactions between brain regions. To evaluate the direction of influences between

390	VMPFC and WM regions, we therefore conducted a DCM analysis (Friston et al.,
391	2003), using DCM12 implemented in SPM12. This analysis was not planned a priori
392	and should therefore be considered exploratory. We focused on the key implication of
393	the PPI results, namely, the possibility that VMPFC exerts a greater effect on WM
394	network regions (here represented by the SPL) under more self-referential conditions.
395	To assess this conjecture more directly, we used the most activated 100 voxels of the
396	VMPFC and bilateral SPLs defined by the group-level self contrast, and saved them
397	as search masks. Then, for each subject, the peak activations within these masks from
398	the first-level analysis were used to create 4-mm-radius-sphere volumes of interest
399	(VOIs), and the activity time series were extracted for each VOI by computing the
400	first eigenvector of all its voxels. These time courses were adjusted for movement
401	parameters and other effects of no interest, while preserving the effects of interest
402	related to the three experimental conditions (Self-Friend, Self-Stranger, and
403	Friend-Stranger).
404	These data were then employed to test a series of models embodying different
405	assumptions about the connectivity and directional influences between the VMPFC
406	and bilateral SPLs. In all models, we assumed intrinsic connections within each
407	region and extrinsic connections between left and right SPL, as well as effects of

408 experimental conditions on each region. Here, to simplify the models, the connection
409 pattern between VMPFC and left SPL was identical to the connection pattern between

410	VMPFC and right SPL. Thus, due to the possible connection patterns between
411	VMPFC and bilateral SPLs, there were four context-independent intrinsic connection
412	matrices (A-matrix): bidirectional connections between VMPFC and SPLs,
413	connection from VMPFC to SPLs, connection from SPLs to VMPFC, and no
414	connection between VMPFC and SPLs. Then, the possible experimental effects on the
415	connection from VMPFC to SPLs and the connection from SPLs to VMPFC were
416	modeled (B-matrix). There was a total of 9 models for each subject, and for each
417	model, we derived the parameters and the free energy, which represents the
418	log-evidence of that model. Then, we compared these models at the group level using
419	random-effects Bayesian model selection (BMS), to identify which model had the
420	highest probability and posterior evidence, and the most probable model was
421	identified according to the exceedance probability (Stephan et al., 2009). The
422	parameter values of the winning model were extracted to assess the difference among
423	conditions using paired t-tests.
424	Stimuli and procedure of tDCS task. Participants in the tDCS study performed a
425	WM task that was identical to the fMRI WM task expect that the duration of the delay
426	period was reduced from 8000 ms to 4000 ms. Prior to performing the WM task,
427	participants were subjected to one of three tDCS regimes. For delivering tDCS, a DC

- 428 Stimulator Plus (NeuroConn) applied a constant current of 1.5 mA for 15 min through
- a pair of electrodes covered in saline-soaked sponges. A  $3 \times 3$  cm<sup>2</sup> forehead electrode 429

430	was located at mid-distance between electrode positions Fz and Fp serving as the
431	stimulating component, and another electrode was placed under the chin as an
432	extracephalic reference. This electrode montage replicated prior studies demonstrating
433	a reliable modulatory effect on hemodynamic responses in VMPFC, maximizing the
434	unipolar stimulation of anterior VMPFC and minimizing the stimulation of other
435	areas (Junghofer et al., 2017; Winker et al., 2018). The forehead electrode was used as
436	the anode to produce excitatory stimulation and as a cathode to produce inhibitory
437	stimulation (Nitsche and Paulus, 2000). Sham stimulation was performed with a
438	current that started out the same as in the anode (or cathode) group but dropped to
439	zero immediately after the initial current injection. The forehead electrode was used
440	anode in half of sham group, and cathode in the other half. To control for possible trait
441	differences in self-prioritization between groups, a measurement of narcissism was
442	conducted for all subjects using the 16-item Narcissistic Personality Inventory
443	(NPI-16) (Ames et al., 2006). There was no difference between the three groups in
444	mean RT, mean accuracy, gender, age, and narcissism score. The NPI-16
445	measurement, associative learning procedure and practice of WM task were
446	performed before the stimulation, and the main WM task was performed immediately
447	after the stimulation phase.

**Results** 

449	Self-associated stimuli are prioritized in WM. Participants were highly accurate on
450	this task, with mean accuracies for the location probe and label probe response being
451	96% and 95%, respectively. Since all participants' mean accuracy was higher than
452	95%, we did not analyze the accuracy data further. Sorted by the type of location
453	probe match response (self-match, friend-match, and stranger-match), RT data were
454	analyzed as a 3-level single-factor within-subjects design. Only correct responses with
455	RTs above 200 ms and within 2.5 standard deviations (SDs) from the subject-specific
456	mean (for each condition) were used for analysis, eliminating less than 1% of trials
457	overall. These trial exclusion criteria were also applied in the subsequent tDCS study.
458	A repeated-measures one-way ANOVA on mean RTs of location probe match trials
459	showed a significant main effect ( $F_{(2, 54)} = 8.72$ , $p = 0.0005$ , $\eta^2 = 0.24$ , see Fig. 2a),
460	with faster responses to self-match trials (755.76 $\pm$ 110.00 ms) than to friend-match
461	trials (776.47 ± 108.82 ms) ( $t_{(27)}$ = 2.36, $p$ = 0.026) and to stranger-match trials
462	$(800.81 \pm 115.24 \text{ ms}) (t_{(27)} = 3.56, p = 0.001)$ , as well as faster responses for
463	friend-match trials than stranger-match trials ( $t_{(27)} = 2.29$ , $p = 0.030$ ). These results
464	successfully replicated those of our previous study (Yin et al., 2019), documenting the
465	prioritization of self-associated stimuli in WM.
466	In the following, we test specific hypotheses of the brain mechanisms mediating
467	this self-prioritization effect. We begin with our first prediction, that the behavioral
468	effect of self-prioritization in WM would be mirrored by enhanced activation for

469	self-associated items in WM regions (in addition to self-referential processing
470	regions), and in more faithful WM representation of the location of self-associated
471	items in visual cortex.
472	Enhanced activation during WM maintenance of self-associated stimuli. We
473	began by characterizing regions involved in WM maintenance, and then assessed their
474	activity profiles as a function of self- versus other-related item maintenance. The
475	different possible combinations of the two memory items resulted in three trial types
476	or pairings: Self-Friend, Self-Stranger and Friend-Stranger. We therefore created a
477	GLM with seven variables, three coding for the delay period for each trial type (our
478	main task phase of interest), three coding for the location probe phase for each trial
479	type, and one coding for the blank screen stage (baseline). To assess general
480	involvement in WM maintenance, we initially contrasted delay period activity
481	(collapsed across conditions) with the blank screen baseline phase (note that neither of
482	these conditions displayed on-screen stimuli).
483	Maintaining WM representations evoked significant activity increases in the
484	supplementary motor area (SMA), bilateral FEF, left IPS, bilateral SPL, bilateral
485	precuneus, and bilateral hippocampus (HIPP) ( $p < 0.001$ , FDR corrected, see Fig. 2b
486	and Table 1 for more details). We next tested whether the prioritization of
487	self-associated items in WM observed in behavior was reflected in activity levels of in
488	WM and self-referential processing regions. To test this hypothesis, we contrasted the

489	condition associated with the most self-referential processing (the Self-Friend
490	condition) with that associated with the least self-referential processing (the
491	Friend-Stranger condition). In contrasting the retention period activity between these
492	two conditions, we found that compared to the Friend-Stranger trials, Self-Friend
493	trials displayed greater activation in the left inferior frontal gyrus (IFG), VMPFC, and
494	bilateral SPL ( $p < 0.05$ , FDR whole-brain corrected, see Fig. 2c and Table 1 for more
495	details). Thus, we observed enhanced activity for maintaining self-associated items in
496	WM in both classic self-referential processing regions (VMPFC) and regions of the
497	WM network (in particular, the SPL). A conjunction analysis formally confirmed the
498	overlap between the self-referential processing effect and WM maintenance related
499	activation in bilateral SPL (see Fig. 3).
500	For illustrative purposes, we extracted the beta-values for each condition from the
501	VMPFC and SPL regions defined by the above-reported contrast, and plotted them in
502	Figure 2d. In addition to recapitulating the results of the ROI-defining contrast (i.e.,
503	greater activity in Self-Friend compared Friend-Stranger trials), these regions also
504	displayed greater activity in the Self-Stranger compared to the Friend-Stranger
505	conditions (VMPFC: ( $t_{(27)} = 3.55$ , $p = 0.001$ ); left SPL: ( $t_{(27)} = 2.25$ , $p = 0.033$ ; the
506	results were equivalent in right SPL), a contrast that is orthogonal to the ROI
507	definition (avoiding circularity). Furthermore, as expected from the WM delay period

508 analysis above, the left SPL exhibited significantly enhanced delay period activity

509	(over baseline) for all three trials types (all $ps < 0.01$ ). Note that VMPFC activity
510	during WM is generally much less pronounced than that in SPL (Fig. 2d). This is
511	expected, as the VMPFC – as part of the DM network – typically exhibits relatively
512	suppressed activation during cognitively demanding tasks like the current one.
513	Importantly, VMPFC shows the greatest release from this relative suppression during
514	the conditions involving the WM maintenance of self-associated cues.
515	In sum, these results showed that in addition to standard WM effects, parietal
516	cortex also displayed a modulation of delay period activity by self-relevance, which
517	was accompanied by typical effects of self-associated items on activity in VMPFC.
518	These findings support one aspect of our first prediction (that is, greater mean activity
519	in WM regions when maintaining self-associated stimuli). We next tested the second
520	aspect, namely that memoranda of self-associated stimuli should be represented more
521	faithfully than those of other-related stimuli, as assessed by decoding success of WM
522	cue locations from delay-period fMRI data.
523	

**Figure 2** 



525

526 Figure 2. Behavioral and neural self-prioritization effects. a Behavioral results from the fMRI WM 527 task replicated previous findings of a self-bias in WM. b Regions showing general involvement in WM 528 maintenance, as defined by enhanced activity during WM delay compared to baseline, include the 529 SMA, bilateral FEF, left IPS, bilateral SPL, bilateral precuneus, and bilateral HIPP (p < 0.001, FDR 530 corrected). c Regions showing enhanced activation during the maintenance of more > less 531 self-associated WM cues include both classic self-referential processing regions (VMPFC) and regions 532 of the WM network (in particular, the SPL). d Beta-values for each condition in VMPFC and left SPL. 533 \* p < 0.05, \*\* p < 0.01, Error bars denote  $\pm 1$  SEM.

### 535 Figure 3



Figure 3. Regions identified by the conjunction GLM analysis. Results showed that left SPL (peak at -27, -72, 57, 32 voxels) and right SPL (peak at 24, -72, 57, 86 voxels) exhibited activation in both the self-referential processing contrast (Self-Friend > Friend-Stranger) and the WM delay-period contrast (delay activity > baseline).

542	Table	1. Activated	brain	regions	in	the	GLM	analy	vsis
				0					/

Contract	Derier	Cluster	Peak	Р	eak M	NI
Contrast	Region	size	t-value	х	у	Z
Self-Friend >	VMPFC	126	5.91	-12	66	-3
Friend-Stranger	L. IFG	76	6.24	-36	27	-9
	L. SPL	75	5.31	-27	-72	57
	R. SPL	140	4.87	24	-75	60
Self-Friend & Self-Stranger	SMA	294	14.30	-3	6	54
& Friend-Stranger > Blank	L. FEF	133	12.61	-36	-3	63
	R. FEF	53	9.88	30	0	66
	R. Hippocampus	53	11.49	21	-39	3
	L. Hippocampus	49	10.31	-18	-42	3
	L. IPS	1037	15.30	-36	-42	42
	L. precuneus	SC	15.05	-15	-72	54
	L. SPL	SC	15.10	-27	-60	54
	R. precuneus	735	12.82	27	-69	54
	R. SPL	SC	12.18	30	-60	54

543 MNI = Montreal Neurological Institute coordinates; L = left; R = right; VMPFC = ventromedial

544 prefrontal cortex; IFG = inferior frontal gyrus; SPL = superior parietal lobule; SMA = supplementary

545 motor area; FEF = frontal eye field; IPS = intraparietal sulcus; SC = same cluster

546

### 547 Enhanced WM representation of self-associated stimuli in visual cortex.

548 According to the "sensory recruitment" view of WM, memoranda should be

549 maintained in relevant sensory cortex, which for the current cue items/locations would

550 be topographically organized, early visual areas. Note that we would not expect to

observe mean (mass-univariate) activity differences between cue conditions, as we are

552 not comparing items for which early visual cortex has differential, selective

- 553 preferences. Rather, in line with previous studies, we reasoned that we should be able
- 554 to decode the locations of cues held in WM from variation in multivoxel activity

555	patterns using MVPA of activity in retinotopically organized visual areas (Sprague et
556	al., 2014; Rahmati et al., 2017; Cai et al., 2019). Importantly, assessing the
557	representations of WM memoranda in visual cortex allowed us to test the second
558	aspect of our first prediction, namely, that the prioritization of self-associated
559	information in WM should be reflected in enhanced neural representation of
560	self-associated locations. To this end, we probed whether the neural classification of
561	self-associated WM cue locations would display higher accuracy than that of
562	others-associated WM cue locations.
563	Recall that in the present task, there were 4 different possible item locations in
564	each visual hemifield (see Fig. 4 and Methods). To define visual areas with reliable
565	retinotopy and sensitivity to stimulation at the WM cue locations, we ran a standard
566	retinotopic localizer (Sereno et al., 1995) and a WM probe location localizer (see
567	Methods). The intersection of visual areas identified by these localizers corresponded
568	to left and right V1, and we employed voxels within this mask for MVPA. To directly
569	compare the neural representation of self-associated locations and other- (that is,
570	friend- or stranger-) associated locations, we only used the trials that involving the
571	self-associated WM cue, and divided these trials into two categories: Self_L and
572	Self_R trials (see Methods).
573	We then trained classifiers on data from left and right V1 at each time point of the
574	WM task trials to decode which of the four possible locations in the contralateral

575	visual hemifield was occupied by the WM cue on a given trial. We ran separate
576	classification analyses for trials where the WM cue in the contralateral hemifield was
577	self-associated or other-associated (see Fig. 4a and Method for more details),
578	resulting in a total of four classifications (left V1 for self-associated cues, left V1 for
579	other-associated cues, right V1 for self-associated cues, right V1 for other-associated
580	cues). For each time point, the classification accuracies of self- (Self_R and Self_L
581	trials) and other-associated cues (Friend and Stranger trials) were averaged.
582	Figure 4b displays the decoding results, plotted as a function of time point (from
583	0 to 18 s). The WM cue location could be decoded at above chance levels 0.25 (all $ps$
584	< 0.001, FWE corrected) for all four classifiers. For comparison, mean
585	mass-univariate activity in this ROI did not differentiate between the three conditions
586	$(F_{(2,50)} = 0.33, p = 0.719, \eta^2 = 0.01)$ . Importantly, as shown in <b>Figure 4b</b> , paired
587	t-tests showed that the classification accuracy for self-associated cues was
588	significantly higher than other-associated cues at the 3rd ( $t_{(25)} = 2.11$ , $p = 0.045$ ), 4th
589	$(t_{(25)} = 4.55, p = 0.0001)$ , 5th $(t_{(25)} = 3.09, p = 0.005)$ , 6th $(t_{(25)} = 3.21, p = 0.004)$ , and
590	7th time point ( $t_{(25)} = 2.48$ , $p = 0.020$ ). Note that, due to hemodynamic lag, the data up
591	until about time points 3 (6s into the delay period) could in principle reflect
592	differential neural responses to the cues themselves, rather than WM maintenance
593	activity. The fact that decoding is successful, and remains superior for self-associated
594	cue locations, over the subsequent time points (up until 14s post cue) shows that this

595	effect clearly extends to activity reflecting WM maintenance per se, however. We next
596	compared the decoding performance of simultaneously maintained self-associated and
597	other-associated cue locations using data averaged over time points 4-6 of the delay
598	period (where decoding was most reliable). Results showed increased decoding
599	accuracy for self-associated cue locations in contralateral visual cortex ( $t_{(25)} = 3.20$ , p
600	= 0.004 for Self_L trials; $t_{(25)} = 2.11$ , $p = 0.045$ for Self_R trials, see <b>Fig. 4c</b> ). These
601	results thus support the idea that the prioritization of self-associated stimuli in WM is
602	reflected in enhanced neural representation of those stimuli in visual cortex.
603	

604 Figure 4





606 Figure 4. Decoding of self- vs. other-associated WM cue locations from early visual cortex. a 607 Examples of self- and other-associated WM cues and V1 areas from a single participant. The left-hand 608 panel depicts the case where a self-associated WM cue is presented in the left visual hemifield and an 609 other-associated cue is presented in the right visual hemifield. The right-hand panel depicts the 610 opposite case.. b Decoding performance of self- and other-associated WM cues displayed as a function 611 of time point. For each time point, the classification accuracies of self- (Self R and Self L trials) and 612 other-associated cues (Friend and Stranger trials) were averaged. The classification accuracy for 613 self-associated cues was significantly higher than for other-associated cues at the 3rd, 4th, 5th, 6th, and 614 7th time point (6-14s post cue). The dashed line shows the chance level (25%). c Decoding 615 performance of simultaneously maintained self-associated and other-associated cue locations (averaged 616 over time points 4-6). The left-hand panel displays the results of the self L trials, and the right-hand 617 panel the results of the self R trials. The vertical axis represents the mean classification accuracy, and 618 the dashed line shows the chance level (25%). L = left, R = right, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001, 619 Error bars denote  $\pm 1$  SEM.

621	In sum, in support of our first prediction, we observed both enhanced activation
622	for maintaining self-associated items in frontoparietal WM regions (in particular the
623	SPL), and more faithful representation of self-associated memoranda in visual cortex.
624	We next turned to our second prediction, namely that the WM self-prioritization effect
625	arises from the influence on WM regions by brain areas specialized for processing
626	self-related information, with the main candidate being the VMPFC region we
627	identified above as displaying greater activation for self- than other-associated items.
628	We first assessed this hypothesis via a functional connectivity analysis and
629	subsequently tested it more rigorously via a tDCS experiment.
630	Self-associated memoranda enhance functional coupling between VMPFC and
631	frontoparietal WM regions. To address the hypothesis that the WM network bias for
632	self-associated cues originates with inputs from brain regions that specialize in
633	self-related processing, we employed a PPI analysis (Friston et al., 1997) to examine
634	changes in the functional coupling (the regression slope of activation) between the
635	VMPFC (the "seed region") and regions in the WM network (the "target regions") as
636	a function of self-associated (Self-Friend) vs. other-associated (Friend-Stranger) WM
637	conditions. While activation in DM regions like the VMPFC typically correlates
637 638	conditions. While activation in DM regions like the VMPFC typically correlates negatively with that in frontoparietal regions subserving top-down attention and WM

640	here predicted the opposite (cf. (Spreng et al., 2010; Gerlach et al., 2011; Dixon et al.,
641	2017)). Specifically, we expected that these self-referential processing regions would
642	exhibit a relative increase in positive functional coupling with WM-related regions
643	during the maintenance of self-associated compared to other-related items, reflecting a
644	biasing of the WM network. The VMPFC seed and WM search space were both
645	defined based on the contrast results reported in the above GLM analysis (Fig. 2). We
646	anticipated that the VMPFC would exhibit increased functional connectivity with
647	WM regions during the maintenance of self- as compared to other-associated
648	locations.
649	In line with this prediction, compared to Friend-Stranger trials, Self-Friend trials
650	showed significantly increased functional connectivity between VMPFC and the
651	SMA, left FEF, and bilateral SPL ( $p < 0.05$ , FDR corrected, see Fig. 5a and Table 2
652	for more details). To directly relate functional coupling to behavior, for each subject,
653	we calculated the behavioral self-prioritization effect by subtracting the self-probe's
654	RT from stranger-probe's RT, and extracted the mean beta-values of the above four
655	WM regions. Then, we conducted a Pearson correlation analysis, which showed that
656	there was a significant positive correlation between the mean increase in connectivity
657	strength and the behavioral self-prioritization effect ( $r = 0.41, p = 0.033$ , see Fig. 5b),
658	thus further corroborating the claim that VMPFC inputs to the WM network mediate
659	the self-prioritization effect.

### 661 Figure 5





663 Figure 5. Functional connectivity (PPI) results. a Regions showing enhanced functional connectivity 664 with the VMPFC (defined by the contrast shown in Fig. 1c) during WM maintenance of 665 self-associated > other-associated memoranda. Enhanced coupling was observed in the SMA, left FEF 666 and bilateral SPL (p < 0.05, FDR corrected). **b** A positive correlation across participants was observed 667 between individual connection strength and behavioral self-prioritization effects. The horizontal axis 668 represents the behavioral self-prioritization effect (defined by subtracting the self-probe's RT from 669 stranger-probe's RT). The vertical axis represents the mean beta-values of the four WM regions. The 670 scatter plot shows the line of best linear fit, and each dot represents data for a single participant.

### 671

Table 2. Brain regions exhibiting enhanced functional coupling in the PPI analysis.

Desien	Cluster	Peak	I	Peak MNI		
Region	size	t-value	х	у	Z	
SMA	52	3.76	-6	9	54	
L. FEF	31	3.87	-27	-3	51	
L. SPL	169	4.80	-18	-75	54	
R. SPL	127	4.31	24	-69	60	

673 MNI = Montreal Neurological Institute coordinates; L = left; R = right; SMA = supplementary motor

674 area; FEF = frontal eye field; SPL = superior parietal lobule

676	Given that PPI analysis does not convey the directionality of influence between
677	brain regions, we followed up the above results with a DCM analysis, geared
678	specifically at probing the interactions between VMPFC and SPL as a function of task
679	conditions (see Methods and Fig. 6a). Note that this analysis was not planned <i>a priori</i> ,
680	and the results should be considered exploratory. We estimated different models of
681	possible influences between these regions and compared their ability to explain the
682	data at the group level using Bayesian model selection. The winning model had an
683	exceedance probability of 0.99, and it included nominally positive (but
684	non-significant) bidirectional intrinsic coupling between VMPFC and SPLs (Fig. 6a)
685	that was modulated by the experimental conditions (Fig. 6b and Table 3). The
686	modulatory effect of task on all three regions' activity was more positive in the
687	Self-Friend than in the Friend-Stranger condition ( $t_{(27)} = 4.03$ , $p = 0.0004$ for left SPL;
688	$t_{(27)} = 3.30, p = 0.003$ for right SPL; $t_{(27)} = 3.68, p = 0.001$ for VMPFC). The
689	task-dependent modulations in reciprocal influence between the VMPFC and SPL
690	were on average inhibitory, but varied by conditions. Specifically, the influence of the
691	VMPFC on processing in the SPL was most inhibitory in the least self-associated WM
692	conditions, as the modulatory effect on the connection from VMPFC to SPL was more
693	negative in Friend-Stranger than in Self-Friend ( $t_{(27)} = 2.11$ , $p = 0.045$ ) and
694	Self-Stranger ( $t_{(27)} = 2.24$ , $p = 0.033$ , see Fig. 6b) conditions. In combination with the
695	PPI results, this could be interpreted as a release from inhibition of the VMPFC on the
696	SPL under conditions of self-associated WM content. By contrast, the coupling from



### 708 Figure 6







714 across-participant means. \* p < 0.05, \*\* p < 0.01.

10 Table 5. Mean (SE) of the modulation parameters for experimental con
---

Conditions	Self-Friend	Self-Stranger	Friend-Stranger
Regions			
L.SPL	0.149 (0.050)	0.100 (0.046)	0.046 (0.061)
R.SPL	0.202 (0.045)	0.162 (0.048)	0.124 (0.046)
VMPFC	0.154 (0.046)	0.125 (0.064)	-0.060 (0.039)
Connections			
L.SPL to VMPFC	-0.232 (0.133)	-0.198 (0.137)	0.118 (0.149)
R.SPL to VMPFC	-0.294 (0.123)	-0.148 (0.118)	-0.144 (0.100)
VMPFC to L.SPL	-0.011 (0.238)	0.009 (0.203)	-0.349 (0.209)
VMPFC to R.SPL	-0.029 (0.211)	-0.025 (0.166)	-0.220 (0.195)

715

### 718 Disrupting VMPFC with cathodal tDCS eliminates the self-prioritization effect

719 in WM. The results of the functional connectivity analysis support the idea that

720 VMPFC was involved in modulating activity in the WM network to favor

721 self-associated items. However, this inference is tentative, as it is based on purely

722 correlational data. In order to test the *necessity* of unperturbed VMPFC function for

the self-bias in WM, we turned to the noninvasive neurostimulation technique of

tDCS, which allows for drawing causal inferences. Specifically, we adopted a tDCS

725 protocol that has recently been shown to reliably modulate VMPFC function

726 (Junghofer et al., 2017; Winker et al., 2018) to perform excitatory (anodal), inhibitory

727 (cathodal), and sham stimulation on this brain region in three independent groups of

728 participants just prior to performing an adapted version of the above WM task (see

729 Methods).

730	A 3 (Group: excitatory, inhibitory, and sham; between-subjects) $\times$ 3
731	(Self-reference: self-match, friend-match, and stranger-match; within-subjects)
732	repeated measures ANOVA showed no main effect of group ( $F_{(2, 87)} = 0.97, p = 0.38$ ,
733	$\eta^2 = 0.02$ ); however, both the main effect of self-reference ( $F_{(2, 174)} = 27.15, p = 5.485$
734	× 10 <sup>-11</sup> , $\eta^2 = 0.24$ ) and the interaction between Group and Self-reference variables ( $F_{(4, -1)}$ )
735	$_{174)} = 3.36$ , $p = 0.011$ , $\eta^2 = 0.07$ ) were significant, with the latter reflecting a
736	differential impact of the stimulation protocols on self-prioritization (see Fig. 7, full
737	behavioral data are shown in Table 4). To elucidate the source of this interaction,
738	separate repeated-measures one-way ANOVAs were conducted in each group. The
739	main effect of self-reference was significant in the anode group ( $F_{(2, 58)} = 17.98, p =$
740	$8.394 \times 10^{-7}$ , $\eta^2 = 0.38$ ) and in the sham group ( $F_{(2, 58)} = 12.89$ , $p = 0.00002$ , $\eta^2 = 0.31$ ),
741	with responses to self-match trials being significantly faster than to both the
742	friend-match trials and stranger-match trials in both groups (all $ps < 0.001$ ). However,
743	the effect of self-reference was abolished in the cathode group ( $F_{(2, 58)} = 1.22, p =$
744	0.301, $\eta^2 = 0.04$ ). Visual inspection of <b>Figure 7</b> might lead one to suspect that this
745	interaction effect was driven by relatively faster responses in friend and stranger trials
746	in the cathode group. To probe this possibility, we performed three one-way
747	between-groups ANOVAs on the RTs of self-match, friend-match, and stranger-match
748	trials, respectively. None of these ANOVAs was significant ( $F_{(2, 87)} = 0.35$ , $p = 0.704$
749	for self-match trials; $F_{(2, 87)} = 1.38$ , $p = 0.257$ for friend-match trials; $F_{(2, 87)} = 1.50$ , $p = 1.50$
750	0.228 for stranger-match trials), indicating that the group x self-reference interaction

- 751 effect was not due to a selective speed-up of the friend and/or stranger conditions in
- the cathode group.
- 753
- 754 Figure 7



Figure 7. Behavioral results on the WM task as a function of tDCS group. A group x self-reference interaction was due to the fact that the effect of self-reference was significant in the anode and sham groups but abolished in the cathode group. \*\*\* p < 0.001, Error bars denote  $\pm 1$  SEM.

760 Table 4. Mean RT (SD) for each group/stimulation condition in the tDCS experiment

	Self	Friend	Stranger	Self-prioritization effect
Anode	696.36 (109.50)	733.63 (105.11)	749.42 (108.37)	53.06 (58.05)
Cathode	693.64 (148.93)	699.63 (148.13)	706.40 (141.19)	12.76 (47.96)
Sham	717.54 (97.59)	751.10 (108.05)	757.83 (117.76)	40.28 (51.42)

761

To directly contrast the self-prioritization effect between groups, we calculated individuals' behavioral self-prioritization effect (subtracting the self-probe's RT from stranger-probe's RT) and compared it between groups. Results showed a significant

765	main effect of Group ( $F_{(2, 87)} = 4.59$ , $p = 0.013$ , $\eta^2 = 0.10$ ), as the self-prioritization
766	effect in cathode group (12.76 $\pm$ 47.96 ms) was significantly weaker than in the anode
767	group (53.06 ± 58.05 ms) ( $t_{(58)}$ = 2.93, $p$ = 0.005) and the sham group (40.28 ± 51.42
768	ms) ( $t_{(58)} = 2.14$ , $p = 0.036$ ). There was no significant enhancement of the
769	self-prioritization effect after anodal compared to sham tDCS, possibly due to a
770	ceiling effect. In conclusion, inhibitory (cathodal) tDCS of VMPFC removed the WM
771	self-prioritization effect, which provides strong support for the hypothesis that
772	VMPFC, well known for its role in self-referential processing, is the source of the
773	self-bias observed in WM.
774	
775	Discussion
776	The present study assessed the neural mechanisms that mediate the prioritization
777	of self-associated information in WM. By pairing a spatial WM task involving self-
778	and other-associated cues with fMRI, we showed that maintaining self- (vs. other-)
779	associated items robustly increased delay-period activity in the VMPFC, as well as in

- 780 components of the WM network, in particular the bilateral SPL. Second, using MVPA,
  - 781 we found that this enhanced activity when maintaining self-associated cues was
- accompanied by a more faithful representation (enhanced decodability) of locations
- 783 corresponding to the self-associated cues in visual cortex. Third, employing PPI
- analysis, we found that individuals' behavioral self-prioritization effect could be

785	accounted for by increased, context-specific functional connectivity between VMPFC
786	and WM-related regions during the maintenance of self-associated cues. DCM
787	indicated a release of a default suppressive influence of VMPFC on SPL under
788	self-associated WM conditions. Finally, we employed tDCS to examine the causal
789	role of the VMPFC in bringing about the WM self-prioritization effect, and found that
790	inhibitory (cathodal) but not anodal or sham stimulation abolished the
791	self-prioritization effect.
792	Our observation of enhanced WM retention period activity in VMPFC and
793	posterior parietal cortex during the maintenance of self-associated stimuli accords
794	well with the prior literature. The VMPFC is perhaps the most frequently implicated
795	region in neuroimaging studies of self-referential processing (Northoff et al., 2006;
796	Lemogne et al., 2012; Murray et al., 2012; Sui et al., 2013), while the SPL is a core
797	component of the WM and dorsal (endogenous) attention networks (Baluch and Itti,
798	2011; Petersen and Posner, 2012; Szczepanski et al., 2013), and has been shown to
799	support the delay period maintenance of WM items in a large number of studies (Todd
800	and Marois, 2004; D'Esposito and Postle, 2015; Rose et al., 2016; Christophel et al.,
801	2017). Note that this parietal focus (and an absence of strong prefrontal involvement)
802	in the current data is likely a consequence of the visuospatial nature of our WM task.
803	Future studies would be required to generalize the current findings to more
804	object-based WM.

805	In the present study, SPL activity was enhanced during the delay-period per se
806	(as in previous work), but it was further enhanced under conditions where
807	self-associated cues had to be maintained. We interpret this activity boost during the
808	maintenance of self-associated cues as reflecting an increased recruitment of
809	top-down attention to support the prioritized WM status of self-associated items.
810	While the detailed neural mechanisms of this prioritization are not yet entirely
811	established, our speculation is concordant with recent resource-based WM accounts.
812	In particular, it has been proposed that WM resources are flexibly (i.e., strategically)
813	distributed among to be maintained items, and that the quality (sharpened
814	representations, as reflected in better decodability) rather than the quantity (e.g., mean
815	neural activity) of WM representations determines performance (Ma et al., 2014;
816	Bays, 2015). Thus, similar to the neural and performance gains observed for
817	retro-cued items in WM (Murray et al., 2013; Myers et al., 2015; Bays and Taylor,
818	2018), we speculate that the self-prioritization effect stems from a biased allocation of
819	internal attention to the self-associated item during WM maintenance.
820	The notion that the increased SPL activity reflects enhanced attentional biasing
821	of WM content is supported by our MVPA findings of more precise delay-period
822	representations of self-associated than other-associated cue locations in visual cortex.
823	While the present paradigm was not optimized to segregate activation associated with
824	the WM encoding vs. maintenance phase, the results suggest strongly that our effects

825	reflect WM maintenance. In particular, due to hemodynamic lag, the BOLD response
826	associated with cue presentation/encoding would be expected to peak around 4-6s into
827	the delay period. Activity related to WM maintenance would be expected to dominate
828	the BOLD response for the subsequent 8s (the duration of the delay period, shifted by
829	the hemodynamic lag), that is, until about 14s after the onset of the delay period. In
830	line with the notion that we are capturing delay-period effects, our time-resolved
831	MVPA results revealed successful cue decoding (and an advantage for self-associated
832	cues) throughout precisely this entire time frame, from 6-14s after delay-period onset
833	(Fig. 4b). Especially the later parts of this phase would clearly not be expected to
834	reflect activity related to initial cue presentation.
835	Prior neuroimaging studies have shown that WM contents can be decoded from
836	multiple regions, ranging from sensory to parietal and prefrontal cortex (Christophel
837	et al., 2012; Emrich et al., 2013; Sreenivasan et al., 2014; Christophel et al., 2017).
838	There is an ongoing debate in the literature whether (frontal and) parietal cortex is
839	directly responsible for representing WM items or whether it supports such
840	maintenance via top-down attentional biasing of sensory cortex (Xu, 2017; Scimeca et
841	
	al., 2018). While the present study was not designed to determine the <i>necessity</i> of
842	al., 2018). While the present study was not designed to determine the <i>necessity</i> of sensory cortex for maintaining WM cue, in line with the sensory recruitment
842 843	al., 2018). While the present study was not designed to determine the <i>necessity</i> of sensory cortex for maintaining WM cue, in line with the sensory recruitment hypothesis (D'Esposito and Postle, 2015; Serences, 2016), we observed clear evidence

845	period. Most importantly for the current purpose, the decoding success for
846	self-associated cue locations was significantly greater than that for (simultaneously
847	presented) other-associated cue locations.
848	What would compel the WM network to prioritize self-associated cue locations
849	in this manner? One can attempt to answer this question at a functional level (why?)
850	and at a mechanistic level (how?). At the functional level, a preference for detecting,
851	encoding, and remembering self-related information could clearly be of benefit to
852	oneself (including at the phylogenetic time scale). Of note, this self-bias appears to be
853	very potent and quasi-automatic: we observed this bias under conditions where we
854	employed meaningless stimuli (colored discs) that were arbitrarily associated with the
855	self or other people, and where self-associated cue locations were no more likely to be
856	probed than other-associated locations. In fact, prior work has shown that this bias
857	even persists when self-associated cues are probed less frequently than
858	other-associated ones, i.e., in situations where the self-bias is clearly not
859	performance-conducive (Sui et al., 2014; Yin et al., 2019).
860	At the mechanistic level, the present study has produced compelling evidence
861	that the neural origin of this bias lies with the VMPFC. First, as expected, the VMPFC
862	exhibited enhanced activity under conditions of self- as compared to other-associated
863	WM maintenance, confirming its prominent role in self-referential processing
864	(Northoff et al., 2006; Qin et al., 2012; Yankouskaya et al., 2017). Second, using PPI

865	analysis, we found that delay-periods where self-associated cues were maintained
866	were characterized by a selective increase in functional connectivity (or a decrease in
867	suppression, as found using DCM) between the VMPFC and regions of the WM
868	network, in particular the SPL. Third, behavioral self-prioritization effects correlated
869	with these PPI context-specific changes in functional coupling across individuals.
870	These results, especially in light of the prior literature implicating the VMPFC in
871	self-referential processing, are strongly suggestive of a biasing influence from the
872	VMPFC on the WM network when self-associated cues had to be maintained. This
873	interpretation is also congruent with previous research reporting increased functional
874	coupling of VMPFC with temporal regions supporting social attention in a task
875	assessing self-bias in a perceptual matching judgement (Sui et al., 2013).
876	Crucially, we tested the above interpretation directly by running a tDCS
877	experiment, adopting a stimulation protocol that has recently been validated as
878	capable of producing distinct modulatory excitatory and inhibitory effects on VMPFC
879	responses, as measured via fMRI (Junghofer et al., 2017; Winker et al., 2018).
880	Whereas groups of participants receiving anodal or sham stimulation displayed the
881	same WM self-bias effect we observed in the fMRI experiment, in the group that
882	received cathodal (inhibitory) stimulation, the self-prioritization effect was
883	completely abolished. This represents causal evidence for the contention that the
884	VMPFC represents the source of the self-focused biasing effects on WM, as

885	anticipated by the above PPI findings. However, as a caveat, it should be noted that
886	we did not directly measure tDCS effects on neural processing in VMPFC in the
887	present experiment. While our behavioral findings are in line with the assumption that
888	the tDCS protocol was successful in modulating VMPFC function, this inference is
889	part reliant on prior studies (Junghofer et al., 2017; Winker et al., 2018), and
890	additional work is still needed to corroborate the possibility of noninvasively
891	influencing self-referential processing in VMPFC. Of note, a within-group
892	experimental design would provide greater sensitivity for assessing such effects.
893	In conclusion, the present study provides novel insights into the brain
894	mechanisms underlying a strong bias for prioritizing the maintenance of
895	self-associated stimuli in WM. Our behavioral, fMRI, and tDCS results provide
896	convergent evidence for the proposal that the VMPFC biases high-level cognitive
897	processing towards self-referential information. In particular, we posit that the
898	VMPFC biases WM representations towards self-associated items via inputs
899	(reflected in enhanced functional coupling) to the WM network (especially posterior
900	parietal cortex), which in turn enhances top-down attentional modulation of sensory
901	regions to emphasize the faithful maintenance of self- (over other-) associated items
902	in memory. Our paradigm and findings provide a unique window into the interaction
903	between social, self-referential processing and high-level cognitive control processes.
904	

## **References**

906 907	Alexopoulos T, Muller D, Ric F, Marendaz C (2012) I, me, mine: Automatic attentional capture by self-related stimuli. Eur J Soc Psychol 42:770-779. <u>https://doi.org/doi:10.1002/ejsp.1882</u>
908 909	Ames DR, Rose P, Anderson CP (2006) The NPI-16 as a short measure of narcissism. J Res Pers 40:440-450. <u>https://doi.org/10.1016/j.jrp.2005.03.002</u>
910	Anticevic A, Repovs G, Shulman GL, Barch DM (2010) When less is more: TPJ and default network
911	deactivation during encoding predicts working memory performance. NeuroImage
912	49:2638-2648. <u>https://doi.org/10.1016/j.neuroimage.2009.11.008</u>
913	Baluch F, Itti L (2011) Mechanisms of top-down attention. Trends Neurosci 34:210-224.
914	https://doi.org/10.1016/j.tins.2011.02.003
915 916	Bays PM (2015) Spikes not slots: noise in neural populations limits working memory. Trends Cogn Sci 19:431-438. <u>https://doi.org/10.1016/j.tics.2015.06.004</u>
917 918	Bays PM, Taylor R (2018) A neural model of retrospective attention in visual working memory. Cogn Psychol 100:43-52. <u>https://doi.org/10.1016/j.cogpsych.2017.12.001</u>
919	Bluhm RL, Clark CR, McFarlane AC, Moores KA, Shaw ME, Lanius RA (2011) Default network
920	connectivity during a working memory task. Hum Brain Mapp 32:1029-1035.
921	<u>https://doi.org/10.1002/hbm.21090</u>
922	Boureau YL, Sokol-Hessner P, Daw ND (2015) Deciding how to decide: Self-control and
923	meta-decision making. Trends Cogn Sci 19:700-710.
924	<u>https://doi.org/10.1016/j.tics.2015.08.013</u>
925	Cai Y, Sheldon AD, Yu Q, Postle BR (2019) Overlapping and distinct contributions of stimulus location
926	and of spatial context to nonspatial visual short-term memory. J Neurophysiol 121:1222-1231.
927	<u>https://doi.org/10.1152/jn.00062.2019</u>
928	Chen AC, Oathes DJ, Chang C, Bradley T, Zhou Z-W, Williams LM, Glover GH, Deisseroth K, Etkin
929	A (2013) Causal interactions between fronto-parietal central executive and default-mode
930	networks in humans. Proc Natl Acad Sci U S A 110:19944-19949.
931	<u>https://doi.org/10.1073/pnas.1311772110</u>
932 933 934	Christophel TB, Hebart MN, Haynes J-D (2012) Decoding the contents of visual short-term memory from human visual and parietal cortex. J Neurosci 32:12983-12989. <u>https://doi.org/10.1523/jneurosci.0184-12.2012</u>
935 936	Christophel TB, Klink PC, Spitzer B, Roelfsema PR, Haynes J-D (2017) The distributed nature of working memory. Trends Cogn Sci 21:111-124. <u>https://doi.org/10.1016/j.tics.2016.12.007</u>
937	D'Esposito M, Postle BR (2015) The cognitive neuroscience of working memory. Annu Rev Psychol
938	66:115-142. https://doi.org/10.1146/annurev-psych-010814-015031

939	Dixon ML, Andrews-Hanna JR, Spreng RN, Irving ZC, Mills C, Girn M, Christoff K (2017)
940	Interactions between the default network and dorsal attention network vary across default
941	subsystems, time, and cognitive states. NeuroImage 147:632-649.
942	<u>https://doi.org/10.1016/j.neuroimage.2016.12.073</u>
943	Emrich SM, Riggall AC, LaRocque JJ, Postle BR (2013) Distributed patterns of activity in sensory
944	cortex reflect the precision of multiple items maintained in visual short-term memory. J
945	Neurosci 33:6516-6523. <u>https://doi.org/10.1523/jneurosci.5732-12.2013</u>
946	Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME (2005) The human brain is
947	intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci U
948	S A 102:9673-9678. <u>https://doi.org/10.1073/pnas.0504136102</u>
949	Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. NeuroImage 19:1273-1302.
950	https://doi.org/10.1016/S1053-8119(03)00202-7
951	Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and
952	modulatory interactions in neuroimaging. NeuroImage 6:218-229.
953	<u>https://doi.org/10.1006/nimg.1997.0291</u>
954 955	Gazzaley A, Nobre AC (2012) Top-down modulation: Bridging selective attention and working memory. Trends Cogn Sci 16:129-135. <u>https://doi.org/10.1016/j.tics.2011.11.014</u>
956	Gerlach KD, Spreng RN, Gilmore AW, Schacter DL (2011) Solving future problems: Default network
957	and executive activity associated with goal-directed mental simulations. NeuroImage
958	55:1816-1824. <u>https://doi.org/10.1016/j.neuroimage.2011.01.030</u>
959	Junghofer M, Winker C, Rehbein MA, Sabatinelli D (2017) Noninvasive stimulation of the
960	ventromedial prefrontal cortex enhances pleasant scene processing. Cereb Cortex
961	27:3449-3456. <u>https://doi.org/10.1093/cercor/bhx073</u>
962	Kesebir S, Oishi S (2010) A spontaneous self-reference effect in memory: Why some birthdays are
963	harder to remember than others. Psychol Sci 21:1525-1531.
964	<u>https://doi.org/10.1177/0956797610383436</u>
965	Lemogne C, Delaveau P, Freton M, Guionnet S, Fossati P (2012) Medial prefrontal cortex and the self
966	in major depression. J Affect Disord 136:e1-e11. <u>https://doi.org/10.1016/j.jad.2010.11.034</u>
967	Liu M, He X, Rotsthein P, Sui J (2016) Dynamically orienting your own face facilitates the automatic
968	attraction of attention. Cogn Neurosci 7:37-44.
969	<u>https://doi.org/10.1080/17588928.2015.1044428</u>
970	Ma WJ, Husain M, Bays PM (2014) Changing concepts of working memory. Nat Neurosci 17:347-356.
971	https://doi.org/10.1038/nn.3655

972	Murray AM, Nobre AC, Clark IA, Cravo AM, Stokes MG (2013) Attention restores discrete items to
973	visual short-term memory. Psychol Sci 24:550-556.
974	<u>https://doi.org/10.1177/0956797612457782</u>
975	Murray RJ, Schaer M, Debbané M (2012) Degrees of separation: A quantitative neuroimaging
976	meta-analysis investigating self-specificity and shared neural activation between self- and
977	other-reflection. Neurosci Biobehav Rev 36:1043-1059.
978	<u>http://dx.doi.org/10.1016/j.neubiorev.2011.12.013</u>
979	Myers NE, Walther L, Wallis G, Stokes MG, Nobre AC (2015) Temporal dynamics of attention during
980	encoding versus maintenance of working memory: complementary views from event-related
981	potentials and alpha-band oscillations. J Cogn Neurosci 27:492-508.
982	<u>https://doi.org/10.1162/jocn_a_00727</u>
983 984 985	Nitsche MA, Paulus W (2000) Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. J Physiol 527:633-639. https://doi.org/10.1111/j.1469-7793.2000.t01-1-00633.x
986	Northoff G, Heinzel A, de Greck M, Bermpohl F, Dobrowolny H, Panksepp J (2006) Self-referential
987	processing in our brain—a meta-analysis of imaging studies on the self. NeuroImage
988	31:440-457. <u>https://doi.org/10.1016/j.neuroimage.2005.12.002</u>
989 990	Petersen SE, Posner MI (2012) The attention system of the human brain: 20 years after. Annu Rev Neurosci 35:73-89. <u>https://doi.org/10.1146/annurev-neuro-062111-150525</u>
991	Qin P, Liu Y, Shi J, Wang Y, Duncan N, Gong Q, Weng X, Northoff G (2012) Dissociation between
992	anterior and posterior cortical regions during self-specificity and familiarity: A combined
993	fMRI–meta-analytic study. Hum Brain Mapp 33:154-164. <u>https://doi.org/10.1002/hbm.21201</u>
994	Rademaker RL, Chunharas C, Serences JT (2019) Coexisting representations of sensory and mnemonic
995	information in human visual cortex. Nat Neurosci 22:1336-1344.
996	<u>https://doi.org/10.1038/s41593-019-0428-x</u>
997 998	Rahmati M, Saber GT, Curtis CE (2017) Population dynamics of early visual cortex during working memory. J Cogn Neurosci 30:219-233. <u>https://doi.org/10.1162/jocn_a_01196</u>
999	Raichle ME (2015) The brain's default mode network. Annu Rev Neurosci 38:433-447.
1000	https://doi.org/10.1146/annurev-neuro-071013-014030
1001	Rose NS, LaRocque JJ, Riggall AC, Gosseries O, Starrett MJ, Meyering EE, Postle BR (2016)
1002	Reactivation of latent working memories with transcranial magnetic stimulation. Science
1003	354:1136-1139. <u>https://doi.org/10.1126/science.aah7011</u>
1004	Schrouff J, Rosa MJ, Rondina JM, Marquand AF, Chu C, Ashburner J, Phillips C, Richiardi J,
1005	Mourão-Miranda J (2013) PRoNTo: Pattern recognition for neuroimaging toolbox.
1006	Neuroinformatics 11:319-337. https://doi.org/10.1007/s12021-013-9178-1

1007	Scimeca JM, Kiyonaga A, D'Esposito M (2018) Reaffirming the sensory recruitment account of
1008	working memory. Trends Cogn Sci 22:190-192. <u>https://doi.org/10.1016/j.tics.2017.12.007</u>
1009 1010	Serences JT (2016) Neural mechanisms of information storage in visual short-term memory. Vision Res 128:53-67. <u>https://doi.org/10.1016/j.visres.2016.09.010</u>
1011	Sereno M, Dale A, Reppas J, Kwong K, Belliveau J, Brady T, Rosen B, Tootell R (1995) Borders of
1012	multiple visual areas in humans revealed by functional magnetic resonance imaging. Science
1013	268:889-893. <u>https://doi.org/10.1126/science.7754376</u>
1014	Sprague Thomas C, Ester Edward F, Serences John T (2014) Reconstructions of information in visual
1015	spatial working memory degrade with memory load. Curr Biol 24:2174-2180.
1016	<u>https://doi.org/10.1016/j.cub.2014.07.066</u>
1017	Spreng RN, Stevens WD, Chamberlain JP, Gilmore AW, Schacter DL (2010) Default network activity,
1018	coupled with the frontoparietal control network, supports goal-directed cognition. NeuroImage
1019	53:303-317. <u>https://doi.org/10.1016/j.neuroimage.2010.06.016</u>
1020 1021	Sreenivasan KK, Curtis CE, D'Esposito M (2014) Revisiting the role of persistent neural activity during working memory. Trends Cogn Sci 18:82-89. <u>https://doi.org/10.1016/j.tics.2013.12.001</u>
1022	Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for
1023	group studies. NeuroImage 46:1004-1017. <u>https://doi.org/10.1016/j.neuroimage.2009.03.025</u>
1024	Sui J, He X, Humphreys GW (2012) Perceptual effects of social salience: evidence from
1025	self-prioritization effects on perceptual matching. J Exp Psychol Hum Percept Perform
1026	38:1105-1117. <u>https://doi.org/10.1037/a0029792</u>
1027	Sui J, Rotshtein P, Humphreys GW (2013) Coupling social attention to the self forms a network for
1028	personal significance. Proc Natl Acad Sci U S A 110:7607-7612.
1029	<u>https://doi.org/10.1073/pnas.1221862110</u>
1030	Sui J, Sun Y, Peng K, Humphreys GW (2014) The automatic and the expected self: separating self- and
1031	familiarity biases effects by manipulating stimulus probability. Atten Percept Psychophys
1032	76:1176-1184. <u>https://doi.org/10.3758/s13414-014-0631-5</u>
1033	Szczepanski SM, Pinsk MA, Douglas MM, Kastner S, Saalmann YB (2013) Functional and structural
1034	architecture of the human dorsal frontoparietal attention network. Proc Natl Acad Sci U S A
1035	110:15806-15811. <u>https://doi.org/10.1073/pnas.1313903110</u>
1036 1037	Todd JJ, Marois R (2004) Capacity limit of visual short-term memory in human posterior parietal cortex. Nature 428:751-754. <u>https://doi.org/10.1038/nature02466</u>
1038	Winker C, Rehbein MA, Sabatinelli D, Dohn M, Maitzen J, Wolters CH, Arolt V, Junghofer M (2018)
1039	Noninvasive stimulation of the ventromedial prefrontal cortex modulates emotional face
1040	processing. NeuroImage 175:388-401. <u>https://doi.org/10.1016/j.neuroimage.2018.03.067</u>

1041	Xu Y (2017) Reevaluating the sensory account of visual working memory storage. Trends Cogn Sci
1042	21:794-815. https://doi.org/10.1016/j.tics.2017.06.013
1043	Yankouskaya A, Humphreys G, Stolte M, Stokes M, Moradi Z, Sui J (2017) An anterior–posterior axis
1044	within the ventromedial prefrontal cortex separates self and reward. Soc Cogn Affect Neurosci
1045	12:1859-1868. <u>https://doi.org/10.1093/scan/nsx112</u>
1046	Yin S, Sui J, Chiu Y-C, Chen A, Egner T (2019) Automatic prioritization of self-referential stimuli in
1047	working memory. Psychol Sci 30:415-423. <u>https://doi.org/10.1177/0956797618818483</u>

а	Day 1	'n	n ~10 min				~60 min					<b>Day 2</b> ~40 min			
	Associati	e learning procedure			WM task t	raining	fMRI measurement of WM task				task	fMRI retinotopic mapping			
	Time $\rightarrow$														
b	Fixation (1 s)	WM cues (1 s)		Delay (8 s)		Location probe (1.5 s) Color-label p				bel probe	obe (1.5 s)				
	+	0 0	+	00	•	C		+	00	朋友 ○ ○	+	00			
	0	0 1		0	2	10	0		1	2		 		18	

а



С

t



y = -69

z = 54







