

---

*Research Articles: Behavioral/Cognitive*

**Behavioral and neural representations of spatial directions across words, schemas, and images**

Steven M. Weisberg<sup>1</sup>, Steven A. Marchette<sup>2</sup> and Anjan Chatterjee<sup>1</sup>

<sup>1</sup>*Department of Neurology, University of Pennsylvania, Philadelphia, PA 19104.*

<sup>2</sup>*Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104.*

DOI: 10.1523/JNEUROSCI.3250-17.2018

Received: 15 November 2017

Revised: 14 March 2018

Accepted: 19 April 2018

Published: 2 May 2018

---

**Author contributions:** S.M.W., S.A.M., and A.C. designed research; S.M.W. performed research; S.M.W. and S.A.M. analyzed data; S.M.W. and A.C. wrote the paper.

**Conflict of Interest:** The authors declare no competing financial interests.

The authors wish to acknowledge NIH grants F32DC015203 to S.M.W., and R01DC012511 and NSF Science of Learning Centers award 1041707 (subcontract 330161-18110-7341) to A.C. They also acknowledge Sam Trinh for stimuli development and pilot testing, Antonio Nicosia for behavioral study data collection, and Russell Epstein for helpful comments and feedback.

Correspondence concerning this article should be addressed to Steven M. Weisberg, at Center for Cognitive Neuroscience, University of Pennsylvania, Philadelphia, PA 19104 USA., Email: [stweis@penncmedicine.upenn.edu](mailto:stweis@penncmedicine.upenn.edu)

**Cite as:** J. Neurosci ; 10.1523/JNEUROSCI.3250-17.2018

**Alerts:** Sign up at [www.jneurosci.org/cgi/alerts](http://www.jneurosci.org/cgi/alerts) to receive customized email alerts when the fully formatted version of this article is published.

Accepted manuscripts are peer-reviewed but have not been through the copyediting, formatting, or proofreading process.

Copyright © 2018 the authors

1 Title: Behavioral and neural representations of spatial directions across words, schemas, and  
2 images

3 Abbreviated Title: Spatial directions in words, schemas, and images  
4

5 Steven M. Weisberg, Steven A. Marchette, & Anjan Chatterjee

6 University of Pennsylvania

7 Steven M. Weisberg, Anjan Chatterjee, Department of Neurology, University of  
8 Pennsylvania, Philadelphia, PA 19104.

9 Steven A. Marchette, Department of Psychology, University of Pennsylvania,  
10 Philadelphia, PA 19104.

11 Correspondence concerning this article should be addressed to Steven M. Weisberg, at  
12 Center for Cognitive Neuroscience, University of Pennsylvania, Philadelphia, PA 19104 USA.

13 Email: [stweis@penncmedicine.upenn.edu](mailto:stweis@penncmedicine.upenn.edu)  
14

15 Number of pages: 41

16 Number of figures: 7

17 Number of tables: 1

18 Number of words for Abstract: 246, Introduction: 650, Discussion: 1364  
19

20 Conflict of Interest Declaration: The authors have no conflicts of interest to declare.

21 Acknowledgements: The authors wish to acknowledge NIH grants F32DC015203 to S.M.W.,  
22 and R01DC012511 and NSF Science of Learning Centers award 1041707 (subcontract 330161-  
23 18110-7341) to A.C. They also acknowledge Sam Trinh for stimuli development and pilot  
24 testing, Antonio Nicosia for behavioral study data collection, and Russell Epstein for helpful  
25 comments and feedback.

26

## Abstract

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

Modern spatial navigation requires fluency with multiple representational formats, including visual scenes, signs, and words. These formats convey different information. Visual scenes are rich and specific, but contain extraneous details. Arrows, as an example of signs, are schematic representations in which the extraneous details are eliminated, but analog spatial properties are preserved. Words eliminate all spatial information and convey spatial directions in a purely abstract form. How does the human brain compute spatial directions within and across these formats? To investigate this question, we conducted two experiments on men and women: a behavioral study that was preregistered, and a neuroimaging study using multivoxel pattern analysis of fMRI data to uncover similarities and differences among representational formats. Participants in the behavioral study viewed spatial directions presented as images, schemas, or words (e.g., "left"), and responded to each trial, indicating whether the spatial direction was the same or different as the one viewed previously. They responded more quickly to schemas and words than images, despite the visual complexity of stimuli being matched. Participants in the fMRI study performed the same task, but responded only to occasional catch trials. Spatial directions in images were decodable in the intraparietal sulcus (IPS) bilaterally, but were not in schemas and words. Spatial directions were also decodable between all three formats. These results suggest that IPS plays a role in calculating spatial directions in visual scenes, but this neural circuitry may be bypassed when the spatial directions are presented as schemas or words.

## 49 Significance Statement

50 Human navigators encounter spatial directions in various formats: words ("turn left"),  
51 schematic signs (an arrow showing a left turn), and visual scenes (a road turning left). The brain  
52 must transform these spatial directions into a plan for action. Here, we investigate similarities  
53 and differences between neural representations of these formats. We found that bilateral  
54 intraparietal sulci represents spatial directions in visual scenes and across the three formats. We  
55 also found that participants respond quickest to schemas, then words, then images, suggesting  
56 that spatial directions in abstract formats are easier to interpret than concrete formats. These  
57 results support a model of spatial direction interpretation in which spatial directions are either  
58 computed for real world action, or computed for efficient visual comparison.

59

## 60 Introduction

61           Humans fluently interpret spatial directions when navigating. But spatial directions are  
62 often presented in distinct representational formats – i.e., words, signs, and scenes – which have  
63 to be converted into a correct series of turns relative to one's facing direction. While spatial  
64 directions conveyed through visual scenes are well-studied, little is known about how spatial  
65 directions are processed within and across these other frequently encountered formats. How does  
66 the human neural architecture for spatial cognition convert spatial maps and scenes into  
67 schematic maps and verbal directions?

68           Representational formats – visual scenes, schematic signs, and words – have distinct  
69 properties, which allow them to convey information differently. Visual scenes convey  
70 navigational information relatively directly – paths are visible – but contain irrelevant  
71 information (e.g., color, objects, context). Words, by contrast, categorize continuously varying  
72 turn angles and, by virtue of being symbolic, are related arbitrarily to the spatial directions  
73 conveyed. Schematic signs, or schemas, are exemplified by arrows in this investigation. Schemas  
74 are simplified visual representations of concepts (Talmy, 2000). Unlike visual scenes, schemas  
75 abstract over properties of spatial directions and omit those that are irrelevant. Unlike words,  
76 schemas maintain an iconic mapping between the spatial direction depicted and its  
77 representational format (i.e., a left-pointing arrow points to the left). Schemas may occupy a  
78 middle ground between images and words representing concepts (for actions, Watson, Cardillo,  
79 Bromberger, & Chatterjee, 2014; for prepositions, Amorapanth et al., 2012; c.f. Gilboa &  
80 Marlatte, 2017).

81           An intricate neural network interprets the spatial content of visual scenes. The occipital  
82 place area (OPA), parahippocampal place area (PPA), retrosplenial complex (RSC), and

83 intraparietal sulcus (IPS) are implicated in different aspects of calculating spatial directions. The  
84 OPA codes egocentric spatial directions (anchored to one's own body position) visible in visual  
85 scenes (Bonner and Epstein, 2017), whereas RSC codes allocentric spatial directions (anchored  
86 to properties of the environment) with respect to a known reference direction (i.e., a major axis  
87 of inside a building, (Marchette, Vass, Ryan, & Epstein, 2015; or north, Vass & Epstein, 2017).  
88 The PPA represents distinct spatial scenes and spatial directions relative to that scene (Epstein,  
89 2008). Unlike OPA, PPA, and RSC, IPS codes egocentric spatial directions that can be either  
90 present in a scene or which were learned and then imagined. For example, Schindler & Bartels  
91 (2013) had participants memorize a circular array of objects and imagine movements with the  
92 same egocentric angle, but anchored to different objects (i.e., "face the lamp, point to the chair"  
93 and "face the chair, point to the vase" would both require a 60° clockwise rotation). Using  
94 multivoxel pattern analysis (MVPA), they showed that IPS exhibited similar patterns of  
95 activation for the same spatial direction. This work used visual scenes to encode spatial  
96 directions. Does IPS represent egocentric spatial directions from arrows (schematic depictions)  
97 and words similarly to visual scenes of egocentric spatial directions?

98         In the current work, we investigate how representational formats affect the behavioral  
99 and neural responses to spatial directions. Our broad hypothesis is that schemas and words elide  
100 the spatial processing required by visual scenes. If true, we predict evidence supporting two  
101 hypotheses: 1) schemas and words are processed more efficiently than scenes, and 2) visual  
102 scenes, but not schemas or words are processed in brain regions known to process spatial  
103 information. To test our first hypothesis, we predict that people will most quickly identify spatial  
104 directions depicted in words ("left" or "sharp right") and schemas (arrows), compared to scenes  
105 (Google Maps images of roads). To test the second hypothesis, in an fMRI study using

106 multivoxel pattern analysis (MVPA), we query the neural representations of spatial directions as  
107 a function of representational format. We expected visual scenes to be processed spatially and  
108 thus spatial directions would be decoded in IPS. We were agnostic if IPS would decode schemas  
109 and words, since these formats are not inherently spatial and need not be processed  
110 egocentrically. We also looked for cross-decoding between all three representational formats.

### 111 **Materials and Methods**

#### 112 **Participants**

113 **Norming Study.** We recruited 42 participants (23 identifying as female) from Amazon  
114 Mechanical Turk. Two participants were removed for responding below chance. Of the  
115 remaining 40 participants (21 identifying as female, 1 did not report gender), 5 participants self-  
116 reported as Asian, 1 as African-American or Black, 2 as Hispanic, 1 as Other, and 29 as White.  
117 Two participants did not report ethnicity. Participants' average age was 34.6 years ( $SD = 12.6$ ).  
118 All but one participant reported speaking English as a first language.

119 **Behavioral Study.** We recruited 48 right-handed participants (27 identifying as female, 1  
120 did not report gender) from a large urban university using an online recruitment tool specific to  
121 that university. 18 participants self-reported as Asian, 13 as African-American or Black, 1 as  
122 American Indian, 5 as Hispanic, 1 as Other, and 9 as Caucasian or White. One participant did not  
123 report ethnicity. Participants' average age was 22.5 years ( $SD = 3.3$ ). Participants reported  
124 speaking English as a first language.

125 **fMRI Study.** We recruited 22 right-handed participants from a large urban university  
126 using an online recruitment tool specific to that university. We excluded data from two  
127 participants because of motion. The resulting sample consisted of 20 participants (11 identifying  
128 as female). 4 participants self-reported as Asian, 4 as African-American or Black, and 12 as

129 Caucasian or White. Participants' average age was 21.4 years ( $SD = 2.7$ ). Participants reported  
130 speaking English as a first language. Laterality quotient indicated participants were right-handed  
131 ( $Min. = 54.17, M = 80.63, SD = 16.41$ ).

### 132 **Experimental materials**

133 **Stimuli.** When given an open number of categories, people freely sort spatial directions  
134 into eight categories (Klippel and Montello, 2007). For the present study, we used seven of those  
135 eight categories: *ahead, left, right, sharp left, sharp right, slight left, and slight right*. We  
136 excluded *behind* because this direction would require the participant to imagine starting at the  
137 top of the image, rather than the bottom.

138 Spatial directions were depicted in three formats – words, schemas, and images. All  
139 stimuli were cropped to be 400x400 pixel squares. For each spatial direction, 24 words were  
140 created in Photoshop by modifying the size (small or large), font (Arial or Times New Roman),  
141 and color (blue, orange, pink, or purple). For each spatial direction, 24 schemas were created in  
142 Photoshop by modifying size (small or large), style (chevron or arrow), and color (blue, orange,  
143 pink, or purple). For the fMRI study, all 24 stimuli were used for each spatial direction. For the  
144 behavioral study, we pseudo-randomly chose three words and schemas to remove (retaining as  
145 close to the same number of colors, sizes, and fonts as possible across the directions), resulting in  
146 21 exemplars per direction.

147 For each spatial direction, 28 images<sup>1</sup> were created from Google Earth. Overhead satellite  
148 views were used to identify roads that turned in that direction, limiting the presence of cars,  
149 arrows on the road, and obscured view (like shadows or trees blocking the view). These 28

---

<sup>1</sup> We are agnostic about whether the images used in this study can actually be considered visual scenes. We use the term "image" below to be consistent with the general terms 'word' and 'schema.' In the introduction and discussion, on the other hand, we discuss 'visual scenes' to connect the domain specific work here with other research on visual scenes. The robust activation of the scene network while participants viewed the images also lead us to speculate that participants treated these stimuli as scenes.

150 images were presented to 40 independent raters on Amazon Mechanical Turk, who answered a  
151 multiple choice question, selecting the spatial direction that best corresponded to the spatial  
152 direction depicted in the image. All raters rated all images. Across each image we compared the  
153 percentage of raters who selected the same direction we chose for each image as a measure of  
154 direction judgment reliability. We selected the top 30 raters on Mechanical Turk (eliminating the  
155 bottom nine participants who rated less than 75% of images correctly). For the fMRI study, we  
156 selected 24 images from each direction to match, as closely as possible, the agreement across  
157 spatial directions (overall agreement = 86.8%,  $SD = 11.6\%$ ). The images differed on overall  
158 agreement across spatial directions,  $F(6, 161) = 3.18, p = .006, \omega^2 = 0.07$ . Ahead images were  
159 the most agreed upon ( $M = 92.6\%$ ) and slight right images ( $M = 80.3\%$ ) were the least agreed  
160 upon. Across spatial directions, the difference in ratings was significant between the slight right  
161 images and the ahead images,  $t(46) = 2.62, p = .012, d = 0.77$ , and between the slight right  
162 images and the left images  $t(46) = 3.173, p = .003, d = 0.94$ . None of the other differences  
163 between pairs of spatial directions reached statistical significance. For the behavioral study, we  
164 used the most agreed upon 21 of these 24 images per spatial direction.

165         We created two versions of words and schemas. For the White background stimuli, words  
166 and schemas were displayed on a white square. For the Scrambled background stimuli, words  
167 and schemas were overlaid on phase-scrambled versions of the images. Two copies of each  
168 phase-scrambled image were used, to provide backgrounds for schemas and words. For  
169 Scrambled backgrounds, spatial directions and backgrounds were randomly paired for each  
170 participant. Sample stimuli (with phase-scrambled backgrounds) can be seen in Figure 1A.

171         **Self-report and debriefing.** All participants in the fMRI and behavioral studies  
172 completed the Verbalizer-Visualizer Questionnaire (VVQ; Kirby, Moore, & Schofield, 1988)

173 before the experimental task. Participants in the fMRI study also completed the Edinburgh  
174 handedness inventory (Oldfield, 1971) before the experimental task. After the experimental task,  
175 participants in the fMRI study completed a debriefing questionnaire, which asked how  
176 participants tried to remember the spatial directions and whether they felt it was difficult to  
177 switch between formats.

178 **Experimental Procedure (Behavioral Study)**

179 To see how efficiently words, schemas, and images are encoded and translated across  
180 formats, participants performed a rolling one-back task. First, participants viewed instructions,  
181 which described the task and included samples of the seven spatial directions in all three formats.  
182 After a brief practice session, participants viewed spatial directions one at a time, responding by  
183 pressing one key to indicate that the current spatial direction was the same as the previously seen  
184 spatial direction, and another key to indicate that the current spatial direction was different from  
185 the previously seen direction. The spatial direction stayed on screen until the participant  
186 responded. Keys ('F' and 'J' on a standard keyboard) were counter-balanced across participants.  
187 Reaction time and accuracy were recorded. We generated a unique continuous carryover  
188 sequence for each participant (Aguirre, 2007) such that each spatial direction and format  
189 appeared before every other format and direction, including itself. This resulted in 441-trial  
190 sequences, of which approximately 1/7 were matches. Half of the participants performed the task  
191 with the White background spatial directions; the other half on the Scrambled background spatial  
192 directions. Because the first trial does not have a previous direction to compare, this trial was  
193 excluded from analysis.

194 **Experimental Procedure (fMRI Study)**

195 To investigate the neural representations of spatial directions across and within formats,  
196 we presented spatial directions one at a time while the participant detected matches and non-  
197 matches in catch trials. Unlike the behavioral study, for the fMRI study we wanted to distinguish  
198 neural activation associated with individual spatial directions, rather than the comparison  
199 between one spatial direction and another. For that reason, participants only responded to  
200 occasional catch trials. The Scrambled spatial directions were used for the fMRI study because a  
201 computational model of early visual cortex processing (the Gist model: Oliva & Torralba, 2006)  
202 could not cross-decode spatial direction across formats when scrambled backgrounds were used,  
203 but could when white backgrounds were used. Using the scrambled backgrounds in the fMRI  
204 study reduced the likelihood of decoding spatial directions across formats because early visual  
205 cortex might be sensitive to low level visual properties that could, for example, distinguish  
206 schemas and images.

207 Continuous carryover sequences (Aguirre, 2007) were generated with 24 conditions -  
208 seven spatial directions in each of the three formats made up 21 conditions; two catch trials (one  
209 match and one non-match); and one null condition. The resulting sequence consisted of 601 trials  
210 (504 spatial directions trials, 48 catch trials, 48 null trials, and the first trial repeated at the end).  
211 Except for the catch trials, which could consist of any stimulus, exemplars were only presented  
212 once per participant. A schematic of the trial structure can be seen in Figure 1B.

213 For spatial direction trials, participants viewed spatial directions one at a time, presented  
214 on screen for 1000 ms with a 2000 ms inter-stimulus interval consisting of a fixation cross.  
215 Participants were instructed to attend to the spatial direction for each trial. On catch trials, a  
216 subsequent spatial direction appeared with the word "Same?" underneath in large red letters. For  
217 these trials, participants pressed one key to indicate that the spatial direction was the same as the

218 one they just saw, and another key to indicate that the spatial direction was different. Keys (the  
219 leftmost and rightmost buttons of a four-button MRI-compatible button box) were counter-  
220 balanced across participants.

221 Catch trials consisted of 1000 ms stimulus presentation, followed by 500 ms fixation,  
222 then 4500 ms of the catch stimulus. The catch stimulus was randomly chosen each time, with the  
223 constraint that it could not be the exact same stimulus. If the catch trial was a match trial, the  
224 spatial direction had to be the same. If the catch trial was a non-match trial, the spatial direction  
225 was randomly chosen from all the other spatial directions. Catch stimuli could be any format.  
226 Null trials consisted of a fixation cross, presented for double the normal trial length, 6000 ms  
227 (Aguirre, 2007).

228 The experimental session was divided into 6 runs. The runs were 100 trials each, except  
229 for the last, which was 101 trials. The second through sixth runs began with the last five trials of  
230 the previous run to re-instate the continuous carryover sequence. These overlap trials, as well as  
231 the catch trials, and null trials, were not analyzed. Because runs contained between 6-9 catch  
232 trials, which were 6000 ms, the runs varied slightly in length, but were approximately 5 min, 50  
233 s. Because of this variation, the scanner collected functional data for 6 min, 12 s. Additional  
234 volumes collected after the stimuli for those trials were finished were discarded. Reaction time  
235 and accuracy were recorded. After each run, the participant received feedback on his  
236 performance (e.g, "You got 6 out of 8 correct.").

237 **MRI Acquisition.** Scanning was performed at the Hospital of the University of  
238 Pennsylvania using a 3T Siemens Trio scanner equipped with a 64-channel head coil. High-  
239 resolution T1-weighted images for anatomical localization were acquired using a three-  
240 dimensional magnetization-prepared rapid acquisition gradient echo pulse sequence [repetition

241 time (TR), 1850 ms; echo time (TE), 3.91 ms; inversion time, 1100 ms; voxel size, 0.9 x 0.9 x 1  
242 mm; matrix size, 240 x 256 x 160]. T2\*-weighted images sensitive to blood oxygenation level-  
243 dependent (BOLD) contrasts were acquired using a multiband gradient echo echoplanar pulse  
244 sequence (TR, 3000 ms; TE, 30 ms; flip angle, 90°; voxel size, 2 x 2 x 2 mm; field of view, 192;  
245 matrix size, 96 x 96 x 80; acceleration factor, 2.). Visual stimuli were displayed by rear-  
246 projecting them onto a Mylar screen at 1024 x 768 pixel resolution with an Epson 8100 3-LCD  
247 projector equipped with a Buhl long-throw lens. Participants viewed the stimuli through a mirror  
248 attached to the head coil.

249       Functional images were corrected for differences in slice timing using FSL slice-time  
250 correction and providing the interleaved slice time order. Images were then realigned to the first  
251 volume of the scan run, and subsequent analyses were performed within the participants' own  
252 space. Motion correction was performed using MCFLIRT (Jenkinson et al., 2002), but motion  
253 outliers were also removed using the Artifact Detection Toolbox  
254 ([http://www.nitrc.org/projects/artifact\\_detect](http://www.nitrc.org/projects/artifact_detect)).

255       For two participants, data from two runs were discarded. For one participant, data were  
256 excluded because the scanning computer crashed during the final run. For a second participant,  
257 data were excluded because performance on the behavioral task was below 50% (chance) for the  
258 final run. All other runs for all other participants exceeded 63% correct.

259       **Multivoxel pattern analysis.** To test whether regions of the brain encoded information  
260 about spatial direction, we calculated, within each participant, the similarities across scan runs  
261 between the multivoxel activity patterns elicited by each spatial direction in each format. If a  
262 region contains information about spatial direction, then patterns corresponding to the same  
263 direction in different scan runs should be more similar than patterns corresponding to different

264 directions (Haxby et al., 2001). Moreover, if this effect is observed for patterns elicited by  
265 stimuli of different formats (i.e., word-schema), then the spatial direction code generalizes across  
266 formats.

267       To define activity patterns, we used general linear models (GLMs), implemented in FSL  
268 (Jenkinson et al., 2012), to estimate the response of each voxel to each stimulus condition (three  
269 formats for each of seven spatial directions) in each scan run. Each runwise GLM included one  
270 regressor for each spatial direction in each format (21 total), regressors for motion parameters,  
271 and nuisance regressors to exclude outlier volumes discovered using the Artifact Detection  
272 Toolbox ([http://www.nitrc.org/projects/artifact\\_detect/](http://www.nitrc.org/projects/artifact_detect/)). Additional nuisance regressors removed  
273 catch trials and the reinstatement trials which began runs 2-5. High-pass filters were used to  
274 remove low temporal frequencies before fitting the GLM, and the first three volumes of each run  
275 were discarded to ensure data quality. Multivoxel patterns were created by concatenating the  
276 estimated responses across all voxels within either the region of interest or the searchlight  
277 sphere. These patterns were then averaged across the first three runs, and then across the second  
278 three runs. For the two participants for whom the final run was discarded, the last two runs were  
279 averaged together.

280       To determine similarities between activity patterns, we calculated Kendall's  $\tau_A$   
281 correlations (Nili et al., 2014) between patterns in the first half and second half scan runs. Before  
282 this computation, we removed the cocktail mean (the average neural activity pattern across all  
283 conditions; Vass and Epstein, 2013) within each format and within each run separately. This  
284 approach normalizes activity patterns across conditions. The pattern of results was unchanged  
285 when the cocktail mean was not removed.

286 We then performed representational similarity analyses by comparing the correlations  
287 between the neural signal across conditions to a theoretical representational dissimilarity matrix  
288 (RDM), which specified how the data would look if a hypothesis were true. This comparison  
289 occurred in one of two ways. 1) If the theoretical RDM was continuous, we correlated the neural  
290 RDM with the theoretical RDM. 2) If the theoretical RDM was binary, we obtained a  
291 discrimination index by averaging a subset of correlations from the neural RDM (e.g., different  
292 direction correlations) and subtracting that from the average of another subset (e.g., same  
293 direction correlations).

294 **Searchlight analysis.** To test for format decoding across the brain, we implemented a  
295 whole-brain searchlight analysis (Kriegeskorte et al., 2006) in which we centered a spherical  
296 ROI (radius, 5 mm) around every voxel of the brain, calculated the spatial direction correlation  
297 within this spherical neighborhood using the method described above, and assigned the resulting  
298 value to the central voxel. Searchlight maps from individual participants were then aligned to the  
299 Montreal Neurological Institute (MNI) template with a linear transformation and submitted to a  
300 second-level random-effects analysis to test the reliability of discrimination across participants.  
301 To find the true type I error rate, we performed Monte Carlo simulations that permuted the sign  
302 of the whole-brain maps from individual participants (Nichols and Holmes, 2002; Winkler et al.,  
303 2014). We performed this procedure 1,000 times across the whole brain. Voxels were considered  
304 significant if they exceeded the  $t$ -statistic of the top 5% of permutations. The mean chance  
305 correlation was 0.

306 **Regions of interest.**

307 **Scene-selective regions.** We identified scene-selective regions of interest (ROIs), Figure  
308 2A-B. These ROIs were defined for each participant individually using a univariate contrast of

309 images>words+schemas, and a group-based anatomical constraint of scene-selective activation  
310 derived from a large number (42) of participants from a previous study (Julian et al., 2012).  
311 Specifically, each ROI was defined as the top 100 voxels in each hemisphere that responded  
312 more to images than to words+schemas and fell within the group-parcel mask for the ROI. To  
313 avoid double-dipping, we defined the ROI using the image>word+schema contrast for one run,  
314 then performed the MVPA analysis as described above on the remaining runs. This method  
315 ensures that all scene-selective ROIs could be defined in both hemispheres in every participant  
316 and that all ROIs contain the same number of voxels, thus facilitating comparisons between  
317 regions.

318 **Visual and Parietal regions.** We defined early visual cortex (EVC) and intraparietal  
319 sulcus (IPS) using the probabilistic atlas from Wang and colleagues (Wang et al., 2015) Figure  
320 2C. These parcels were registered to participants'-own-space and voxels were extracted. The  
321 MVPA analysis was then performed as described above using all data. We analyzed all IPS  
322 regions in one combined ROI (the union of all voxels from the Wang et al IPS parcels) without  
323 further selection of voxels from functional comparisons.

#### 324 **Experimental Design and Statistical Analysis**

325 We conducted 2 experimental studies. The sample size for the behavioral study was  
326 selected based on a power analysis from a smaller pilot study with 15 participants. The  
327 behavioral study and reported analyses were preregistered on Open Science Framework  
328 (<https://osf.io/5dk37/>), but the code used to analyze the data was altered, because of software  
329 bugs, which were unknown at the time of the original registration. Within- and between-subject  
330 factors and materials for that study can be found on Open Science Framework, and in the method  
331 section above.

332 The sample size for the imaging study was based on previous similar studies (Schindler  
333 and Bartels, 2013; Marchette et al., 2015) that examined within-subject differences in MVPA of  
334 the BOLD fMRI signal. Although we looked at individual differences in an exploratory fashion,  
335 we interpret these results with caution. Details and important parameters for the imaging study  
336 can be found in the Method section.

337 Across both studies, where appropriate, we corrected for multiple comparisons and report  
338 in the text how these determinations were made.

### 339 **Results**

#### 340 **Behavioral Study**

341 Accuracy on the rolling one-back task was high ( $M = 93.7\%$ ,  $SD = 21.57$ , Range =  
342 [75.0% - 99.8%]). Including incorrect trials, reaction time overall was  $M = 1.31s$ ,  $SD = 0.34s$ ,  
343 Range = [0.70s - 2.14s]. We excluded data from one participant because of low accuracy  
344 (53.9%) and fast reaction time (0.24s) compared to data from the rest of the sample.

345 Participants did not differ in accuracy when responding to schemas ( $M = 94.7\%$ ,  $SD =$   
346 4.2%), and words ( $M = 94.4\%$ ,  $SD = 4.2\%$ ),  $t(46) = 1.21$ ,  $p = .23$ ,  $d = 0.14$ , but were more  
347 accurate responding to schemas compared to images ( $M = 91.3\%$ ,  $SD = 7.2\%$ ),  $t(46) = 5.26$ ,  $p =$   
348 .0000004,  $d = 0.97$ . Participants were also more accurate for words compared to images,  $t(46) =$   
349 4.68,  $p = .000003$ ,  $d = 0.88$ . This result shows that, compared to schemas and words, the spatial  
350 directions in the images were more difficult to identify (i.e., it was possible to interpret a slight  
351 right turn as a right or an ahead). To avoid this confound and a speed-accuracy tradeoff, we  
352 excluded incorrect trials and only analyzed reaction times for correct trials across formats.

353 **Schemas are processed more quickly than images or words.** In addition to excluding  
354 correct trials, we excluded trials for which the participant responded especially slowly – greater

355 than two standard deviations above his/her mean reaction time. We also excluded trials for which  
356 the answer was "same," because these trials occurred relatively infrequently and could be  
357 considered oddball trials. They also required a different response than the other trials. All further  
358 analyses exclude trials as described above.

359 Figure 3 displays the main reaction time results for the behavioral study. Reaction time  
360 for schemas was quicker ( $M = 1.13\text{s}$ ,  $SD = 0.28\text{s}$ ) than for images ( $M = 1.27\text{s}$ ,  $SD = 0.35\text{s}$ ),  $t(46)$   
361  $= 7.40$ ,  $p = .000000002$ ,  $d = 1.19$ , and words ( $M = 1.18\text{s}$ ,  $SD = 0.26\text{s}$ ),  $t(46) = 3.09$ ,  $p = .003$ ,  $d =$   
362  $0.49$ . Reaction time for words was also quicker than for images,  $t(46) = 4.38$ ,  $p = .00007$ ,  $d =$   
363  $0.97$ .

364 **Same-format advantage.** Comparing spatial directions was faster when the preceding  
365 stimulus was in the same format. We calculated the average reaction time for each current format  
366 (the trial for which a response is generated) separately based on whether the previous trial was  
367 the same or a different format. The same-format advantage is operationalized as the difference in  
368 reaction time between same-format-preceding trials and different-format-preceding trials. Higher  
369 numbers indicate faster responses for same-format comparisons than different-format  
370 comparisons. Images ( $M = 0.06\text{s}$ ,  $SD = 0.16\text{s}$ ), one-sample  $t(46) = 2.47$ ,  $p = .017$ ,  $d = 0.38$ ,  
371 schemas ( $M = 0.15\text{s}$ ,  $SD = 0.15\text{s}$ ), one-sample  $t(46) = 6.94$ ,  $p = .00000001$ ,  $d = 1.00$ , and words  
372 ( $M = 0.12\text{s}$ ,  $SD = 0.13\text{s}$ ), one-sample  $t(46) = 6.38$ ,  $p = .00000008$ ,  $d = 0.92$ , all showed  
373 significant same-format advantages. Comparing the same-format advantage between images,  
374 schemas, and words revealed that schemas showed a larger same-format advantage than images,  
375  $t(46) = 3.43$ ,  $p = .001$ ,  $d = 0.50$ , and a marginally larger advantage than words,  $t(46) = 1.86$ ,  $p =$   
376  $.07$ ,  $d = 0.34$ . Words showed a significantly larger same-format advantage than images,  $t(46) =$

377 2.35,  $p = .02$ ,  $d = 0.28$ . Comparing spatial directions was always faster when these comparisons  
378 were within format, but schemas showed this effect most strongly.

379 **Phase-scrambled backgrounds did not show behavioral effects.** As reported in the  
380 Method section, the Gist model could not decode spatial directions across formats when  
381 scrambled backgrounds were used for schemas and words, but could do so with white  
382 backgrounds. Despite this finding, we did not find behavioral effects based on background.  
383 Reaction time (on all trials) was similar for White ( $M = 1.28s$ ,  $SD = 0.32s$ ) and Scrambled  
384 backgrounds ( $M = 1.33s$ ,  $SD = 0.36s$ ),  $t(45) = 0.43$ ,  $p = .67$ ,  $d = 0.15$ . Accuracy was similar for  
385 White ( $M = 93.1%$ ,  $SD = 5.50%$ ) and Scrambled backgrounds ( $M = 94.2%$ ,  $SD = 4.30%$ ),  $t(45) =$   
386  $0.76$ ,  $p = .45$ ,  $d = 0.22$ . In addition, none of the above analyses interacted with the background  
387 condition.

388 **Reaction time correlates with egocentric not visual angular distance between trials.**  
389 We instructed participants to imagine the directions as egocentric, with respect to their own body  
390 position, but wondered whether reaction time data were consistent with participants following  
391 this instruction. Thus, we calculated the angular distance between each pair of trials in two ways.  
392 The visual angle was calculated as the absolute value of the angular distance between the current  
393 and previous trials. The egocentric angle was calculated similarly, except that all angular  
394 distances were calculated as if sharp right and sharp left were maximally far apart. We called this  
395 the egocentric angle because relative to one's facing direction, the head cannot rotate behind the  
396 body, thus this angle calculation preserves egocentric validity. For example, the angular distance  
397 between sharp right and sharp left was  $90^\circ$  for visual angle, but  $270^\circ$  for egocentric angle. To  
398 determine whether there was a significant correlation within participants, we calculated the  
399 Pearson's correlation between each participant's reaction time on that trial with the visual and

400 egocentric angular distance between that trial and the previous trial. We then conducted one-  
401 sample *t*-tests to determine if there was a significant correlation in our sample, and within-  
402 subject *t*-tests to compare correlations. We found that egocentric angles correlated with reaction  
403 time positively ( $M = .040$ ,  $SD = .065$ ),  $t(46) = 4.21$ ,  $p = .0001$ ,  $d = 0.61$ , but visual angles did not  
404 ( $M = .0074$ ,  $SD = .064$ ),  $t(46) = 0.79$ ,  $p = .43$ ,  $d = 0.12$ . These patterns significantly differed from  
405 each other,  $t(46) = 2.75$ ,  $p = .009$ ,  $d = 0.82$ . This pattern of results was obtained within each  
406 format separately, and angular distance correlation did not interact with format. This pattern of  
407 results reveals that participants interpreted spatial directions egocentrically because longer  
408 reaction times were associated with larger egocentric but not visual angle distances.

409 **Individual differences in cognitive style did not correlate with reaction time.** No  
410 measures of reaction time correlated with either measure of the Verbalizer-Visualizer  
411 Questionnaire, all  $p$ 's  $> .08$ .

#### 412 **fMRI Study**

413 **Behavioral performance during the fMRI task.** Responses to the catch trials during the  
414 fMRI task were accurate ( $M = 89.9\%$ ,  $SD = 6.74\%$ ). Behavioral responses during one run for one  
415 participant fell below chance (43%, 3/7 correct). fMRI data for that run for that participant were  
416 excluded.

#### 417 **Spatial direction decoding.**

418 ***Within-format decoding of spatial direction in ROIs.*** Figure 4A displays the within-  
419 format contrasts used to calculate whether spatial directions were decoded in each ROI. The  
420 whole grid represents a theoretical RDM of three separate contrasts: Same minus different spatial  
421 direction within each format. These contrasts were performed on each participant's neural RDM,  
422 calculated as described in the Method section by correlating the averaged parameter estimates for

423 each trial type (e.g., a slight right word, or a sharp left schema) separately for the first and second  
424 half of each participant's runs. Separately for each format, grey squares were subtracted from  
425 colored squares. White squares were omitted.

426 Within-format results for each of the ROIs are displayed in Figure 4B. Spatial direction  
427 was decoded within images in IPS ( $M = 0.08$ ,  $SD = 0.10$ ), one-sample  $t(19) = 3.56$ ,  $p = .002$ ,  $d =$   
428  $0.80$ . EVC did not decode spatial direction within images, ( $M = 0.01$ ,  $SD = 0.09$ ), one-sample  
429  $t(19) = 0.73$ ,  $p = .47$ ,  $d = 0.10$ . IPS decoded spatial direction within images significantly more  
430 than EVC,  $t(19) = 2.97$ ,  $p = .0078$ ,  $d = 0.69$ . Scene regions did not decode spatial directions  
431 within images ( $M_{OPA} = -0.003$ ,  $SD_{OPA} = 0.04$ ,  $t_{OPA}(19) = 0.36$ ,  $p = .72$ ,  $d = -0.08$ ;  $M_{PPA} = -0.006$ ,  
432  $SD_{PPA} = 0.03$ ,  $t_{PPA}(19) = 0.90$ ,  $p = .38$ ,  $d = -0.20$ ;  $M_{RSC} = -0.003$ ,  $SD_{RSC} = 0.04$ ,  $t_{RSC}(19) = 0.34$ ,  
433  $p = .74$ ,  $d = -0.08$ ).

434 Spatial directions were not decoded for schemas or for words in any of the ROIs (all  $p$ 's  $>$   
435  $.26$ ).

436 ***Cross-format decoding of spatial direction in ROIs.*** We also wished to learn if spatial  
437 directions could be decoded independently of the visual properties of individual formats. A brain  
438 region would show evidence of cross-format decoding of spatial direction if the correlation  
439 between the same spatial direction, presented in different formats, exceeded the correlation  
440 between different spatial directions, presented in different formats. Figure 5A displays the cross-  
441 format decoding theoretical RDM. Grey squares (different direction, different format) are  
442 subtracted from black squares (same direction, different format) to yield the degree of  
443 generalization. White squares are omitted.

444 Results for each of the ROIs are displayed in Figure 5B. Cross-format spatial directions  
445 were decoded in IPS ( $M = 0.04$ ,  $SD = 0.06$ ), one-sample  $t(19) = 2.64$ ,  $p = .0128$ ,  $d = 0.67$ . There

446 was marginally significant cross-format decoding in EVC ( $M = 0.01$ ,  $SD = 0.02$ ), one-sample  
447  $t(19) = 1.80$ ,  $p = .087$ ,  $d = 0.50$ . IPS decoded spatial directions across formats marginally more  
448 than EVC,  $t(19) = 2.09$ ,  $p = .051$ ,  $d = 0.66$ . The scene regions did not decode spatial directions  
449 across formats ( $M_{OPA} = -0.003$ ,  $SD_{OPA} = 0.02$ ,  $t_{OPA}(19) = 0.79$ ,  $p = .44$ ,  $d = -0.15$ ;  $M_{PPA} = 0.004$ ,  
450  $SD_{PPA} = 0.015$ ,  $t_{PPA}(19) = 1.21$ ,  $p = .24$ ,  $d = 0.27$ ;  $M_{RSC} = 0.0009$ ,  $SD_{RSC} = 0.01$ ,  $t_{RSC}(19) = 0.34$ ,  
451  $p = .74$ ,  $d = 0.09$ ). These data reveal that IPS contain cross-format representations of spatial  
452 direction, but EVC and scene regions do not.

453         Within IPS we wanted to know whether cross-format decoding of spatial direction was  
454 driven by particular pairs of formats. For example, it is possible that spatial direction decoding  
455 was high between images and schemas, but comparatively lower between images and words. To  
456 investigate this, we conducted follow-up contrasts between each pair of formats similar to the  
457 omnibus test above (e.g., same direction, different format minus different direction, different  
458 format for images versus schemas). These follow-up contrasts revealed significant schema-word  
459 decoding ( $M = 0.036$ ,  $SD = 0.07$ ), one-sample  $t(19) = 2.23$ ,  $p = .04$ ,  $d = 0.51$  and marginally  
460 significant image-schema decoding ( $M = 0.024$ ,  $SD = 0.05$ ), one-sample  $t(19) = 1.99$ ,  $p = .06$ ,  $d$   
461  $= 0.48$ . Image-word decoding was not significant ( $M = 0.008$ ,  $SD = 0.07$ ), one-sample  $t(19) =$   
462  $0.52$ ,  $p = .61$ ,  $d = 0.11$ . These follow-up contrasts were not significantly different from each  
463 other (all pairwise  $p$ 's  $> .25$ ). This pattern of results suggest that schemas may occupy an  
464 intermediary role, sharing neural responses in IPS with images and words respectively in a way  
465 not seen with images and words.

466         *Spatial direction decoding in searchlights.* Although we had specific predictions about  
467 regions of the brain that might decode spatial directions, we also conducted exploratory analyses  
468 to assess within- and cross-format spatial direction decoding at the whole-brain level. We did so

469 to see whether any regions of the brain outside IPS decoded spatial directions in words or  
470 schemas. None of these analyses survived correction for multiple comparisons. So we report our  
471 observations at a lower threshold and advise caution in interpretation. We report results that  
472 exceed a lower threshold ( $p < .0005$ , uncorrected). Within-format decoding of images occurred  
473 in left posterior parietal cortex, extending into left medial parietal cortex – a region consistent  
474 with our IPS ROI, as well as a left lateral frontal region. Within-format decoding of schemas  
475 occurred in left premotor cortex, and within-format decoding of words occurred in a small region  
476 in the brain stem. Cross-format decoding revealed a small region near left visual area MT.

477       **Spatial direction similarity analysis.** The preceding analyses reveal that IPS can  
478 distinguish between the seven spatial directions within images, and across formats. There are two  
479 possible ways IPS could do this. IPS could be creating seven arbitrary and ad hoc categories for  
480 each spatial direction, which could allow any type of information to be decoded. If this  
481 interpretation is correct, the IPS' role in spatial direction coding would be that it is creating a  
482 problem space onto which any possible stimulus categories could be mapped. For example, if the  
483 task were to sort stimuli based on seven colors, IPS would create seven color categories, which  
484 would be most similar to themselves (e.g., red is most similar to red), and different from all  
485 others. On the other hand, IPS could be involved because it helps distinguish spatial directions,  
486 specifically. If this interpretation is correct, the IPS' role in spatial direction coding would be that  
487 it constructs a spatial representation of the possible directions. A counter-example for color  
488 would be that IPS contains a color-wheel representation. To distinguish which of these  
489 possibilities is correct, we can analyze off-diagonal spatial direction similarity. We would expect  
490 categories of turns (e.g., left to slight left) to be more similar to each other than to more distant  
491 turns (e.g., left to sharp right). We created a new theoretical RDM in which all left turns (sharp

492 left, left, and slight left) were similar to each other, and dissimilar to all right turns (and vice  
493 versa for right to left turns). Ahead directions were coded as dissimilar from everything else. We  
494 excluded the diagonal to ensure that these results are not recapitulations of the spatial direction  
495 decoding analyses above. That is, this analysis captures similarity among non-identical spatial  
496 directions to show that IPS neural patterns contain spatial information (not arbitrary category  
497 information).

498         We found that the neural pattern of activity in IPS in response to images correlated more  
499 strongly between left turns than across left and right turns ( $M = 0.036$ ,  $SD = 0.063$ ),  $t(19) = 2.59$ ,  
500  $p = .018$ ,  $d = 0.57$ . This pattern was not the case for schemas, ( $M = -0.018$ ,  $SD = 0.086$ ),  $t(19) =$   
501  $0.09$ ,  $p = .93$ ,  $d = -0.21$ , nor for words, ( $M = -0.017$ ,  $SD = 0.075$ ),  $t(19) = 1.02$ ,  $p = .32$ ,  $d = -0.23$ ,  
502 nor across formats, ( $M = 0.009$ ,  $SD = 0.037$ ),  $t(19) = 1.10$ ,  $p = .29$ ,  $d = 0.24$ . This result provides  
503 evidence that images were represented spatially, by distinguishing left from right turns, in IPS,  
504 and not as seven arbitrary and ad hoc categories. Although this analysis shows that IPS codes  
505 spatial content, the theoretical RDM we chose was not the only possible one. We also conducted  
506 a representational similarity analysis wherein we correlated the neural RDM with a spatial  
507 direction model where similarity linearly decreased as a function of spatial angle, but this  
508 analysis did not achieve statistical significance. We thus interpret this result as evidence of  
509 spatial content in IPS, but do not feel strongly that the representation is categorical (i.e., all lefts  
510 are more similar to each other than to rights).

511         **Format decoding.** In the following analyses, we removed the cocktail mean within run,  
512 across all formats.

513         **Format decoding in ROIs.** In addition to direction coding, we wanted to determine  
514 whether the format of stimuli was represented in these ROIs. The theoretical RDM for this

515 contrast is presented in Figure 6A. For this analysis, we excluded correlations between stimuli  
516 that were the same direction and the same format (white squares in Figure 6A). To decode  
517 format, a region would show higher correlations between stimuli that were the same format  
518 compared to stimuli that were different formats (black squares minus grey squares in Figure 6A).

519 Results from the omnibus format decoding contrast can be seen in Figure 6B. Format  
520 could be decoded in IPS ( $M = 0.14$ ,  $SD = 0.09$ ), one-sample  $t(19) = 7.25$ ,  $p = .0000007$ ,  $d = 1.56$ ,  
521 and EVC ( $M = 0.02$ ,  $SD = 0.03$ ), one-sample  $t(19) = 3.58$ ,  $p = .002$ ,  $d = 0.67$ , although format  
522 decoding was significantly higher in IPS than EVC  $t(19) = 6.39$ ,  $p = .000004$ ,  $d = 1.63$ . OPA ( $M$   
523  $= 0.04$ ,  $SD = 0.03$ ), one-sample  $t(19) = 7.07$ ,  $p = .000001$ ,  $d = 1.33$ , and PPA ( $M = 0.008$ ,  $SD =$   
524  $0.01$ ), one-sample  $t(19) = 2.54$ ,  $p = .02$ ,  $d = 0.80$ , also decoded format, although RSC did not ( $M$   
525  $= 0.003$ ,  $SD = 0.008$ ), one-sample  $t(19) = 1.39$ ,  $p = .18$ ,  $d = 0.38$ .

526 We wanted to know whether the regions that significantly decoded format generally (IPS,  
527 EVC, OPA, and PPA) could decode pairwise formats. We thus looked at schema-word, schema-  
528 image, and image-word decoding separately for each ROI. See Table 1 for the complete results.  
529 In sum, pairwise formats could be decoded to some extent in each ROI except RSC. In IPS and  
530 OPA, all three pairs of formats could be distinguished, whereas PPA predominantly dissociated  
531 images from the schemas and words.

532 To visualize whether format decoding was similar across IPS and OPA, each ROI's  
533 neural RDM was submitted to multidimensional scaling (MDS), which are projected into two-  
534 dimensional maps of each spatial direction and format. In these maps (Figure 6C), each arrow  
535 depicts one trial type, and the distance between arrows can be interpreted as the pair's  
536 representational dissimilarity. For ease of interpretation, and to be consistent with the spatial  
537 decoding in IPS described in the spatial direction similarity analysis, we collapsed across left,

538 right, and ahead. The MDS plots emphasize that while both regions distinguish between all three  
539 formats, schemas and words are more clearly disambiguated in IPS. Notably, format accounts for  
540 a large proportion of the variance captured by both regions, in spite of the fact that participants  
541 were asked to respond only to the spatial direction in the stimulus independent of the format.

542 **Format decoding in searchlight analyses.** Format decoding was robust within our  
543 regions of interest. We also queried the whole brain. We ran two searchlight analyses to see  
544 where formats were decoded across the whole brain. First, we analyzed which regions  
545 represented images as more similar to images than images to schemas or words. These regions  
546 are visualized in hot colors in Figure 7. In addition to parietal lobes, canonical scene regions  
547 (OPA, RSC, PPA) have higher correlations between images than with images to other formats.  
548 Second, we analyzed which regions represented schemas as more similar to schemas compared  
549 to words, and words more similar to words than schemas. This analysis uses the same baseline,  
550 word-schema correlations, and thus cannot distinguish whether these regions represent words as  
551 more similar to words, schemas more similar to schemas, or both. These regions are visualized in  
552 cool colors in Figure 7. Here, we saw bilateral fusiform gyrus, and inferior lateral occipital  
553 cortex, regions which have been implicated in word and object processing.

554 **Individual differences in cognitive style.** In an exploratory analysis, we correlated both  
555 dimensions of the Verbalizer-Visualizer Questionnaire (VVQ) with spatial direction decoding,  
556 within and across formats, in IPS. Correcting for multiple comparisons, no significant spatial  
557 direction decoding correlations were found. This question would be addressed more  
558 appropriately with a larger sample size. Responses to the de-briefing questionnaire also indicated  
559 that some participants preferred to say words to themselves, whereas others preferred to picture  
560 directions, or even imagine part of their body (e.g., their left shoulder for slight left). This

561 variability in self-reported strategy suggests individual differences in cross-format decoding that  
562 a higher-powered study could address.

563 **Discussion**

564 We aimed to investigate how spatial directions conveyed by distinct representational  
565 formats – visual scenes, schemas, and words – are behaviorally processed and neurally  
566 organized. We hypothesized that schemas and words elide spatial processing required by visual  
567 scenes and are processed more efficiently. This work bridges non-human models of navigation  
568 and cognitive mapping from visual scenes (Poucet, 1993; Etienne et al., 1996; Chen et al., 2013),  
569 with human research, which can investigate schematic and verbal communication of spatial  
570 directions.

571 Our findings support a model of spatial direction processing which taps a network that  
572 computes paths in visual scenes (Schindler & Bartels, 2013), but eschews in-depth spatial  
573 computations for efficient format-specific visual processing. Computing spatial directions from  
574 visual scenes requires imagining travel on paths shown. Computing spatial directions from words  
575 and schemas requires only visual identification. Visual scenes contain concrete detail, irrelevant  
576 to the spatial direction, but allow navigators to imagine traveling through the scene. By contrast,  
577 schemas and words contain easily distinguishable abstract direction information, but do not  
578 invite imagining travel in the same way as scenes.

579 In support of this model, we report three main findings. First, people responded to  
580 schemas and words more quickly than to scenes. Second, the intraparietal sulcus (IPS) bilaterally  
581 decoded spatial directions in scenes (and the decoding was structured into spatial categories,  
582 rather than arbitrary categories), and across the three formats, but not within schemas or words.  
583 These two findings suggest that, compared to words and schemas, scenes require relatively

584 costly spatial computation to decode spatial directions in IPS. Third, format decoding  
585 independent of spatial directions was robust in ROI and whole-brain searchlight analyses. This  
586 finding suggests that, despite being task irrelevant (i.e., once the spatial direction is encoded,  
587 participants are better off discarding format, since spatial direction can be queried in any format),  
588 formats tap distinct neural pathways to convey relevant information.

589         Why might scenes be processed more slowly than schemas and words? First, unlike  
590 schemas, which discard irrelevant visual information and distill conceptually-important content,  
591 visual scenes contain detail unnecessary to compute the direction being depicted. Second,  
592 directions conveyed by schemas (at least in the current experiment) and words contain the exact  
593 same information about the spatial direction. Visual scenes can deviate from, for example, an  
594 exact 90° left turn. Thus, the direction in a visual scene must be computed for each presentation,  
595 then compared to the previous stimulus, whereas schemas and words need not be processed with  
596 this level of discrimination.

597         If spatial directions are computed from visual scenes, brain regions which support  
598 direction processing should contain representations of spatial directions for visual scenes, but not  
599 for schemas or words. This pattern was observed in the IPS bilaterally, regions of the brain  
600 implicated in egocentric spatial direction processing (Karnath, 1997; Whitlock et al., 2008;  
601 Galati et al., 2010; Schindler and Bartels, 2013).

602         We also found cross-decoding between schemas, words, and visual scenes in the IPS  
603 bilaterally. One explanation of our results is that when an individual views a scene, the IPS  
604 compute egocentric spatial directions from visual scenes by imagining the path of travel,  
605 resulting in a strong signal for each direction. However, when an individual views a schema or  
606 word, discerning spatial direction does not require IPS to compute egocentric spatial directions,

607 yet it does so transiently, resulting in a weak signal. Within schema and word formats, this weak  
608 signal might not itself be decodable. Comparing the weak signal from schemas to the strong  
609 signal from scenes could yield cross-format decoding.

610 We did not observe spatial direction decoding in OPA, PPA, or RSC. The current results  
611 are not necessarily at odds with previous research showing spatial direction decoding in OPA  
612 (Julian et al., 2016; Bonner and Epstein, 2017) because our participants did not view walkable  
613 pathways. The OPA is causally involved in representing spatial directions defined by visual  
614 scene geometry and boundaries (e.g., constrained by hallways, counters, etc.) for obstacle  
615 avoidance. The lack of spatial direction decoding in PPA aligns with the hypothesis that the PPA  
616 codes a local visual scene (Epstein, 2008) in a viewpoint-invariant manner (Epstein et al., 2003).  
617 This hypothesis is supported by data showing that the neural representation in PPA is consistent  
618 across the same viewpoint of the same scene, but different when the viewpoint changes,  
619 suggesting that the PPA encodes spatial direction only with respect to the same scene. PPA may  
620 not have decoded directions in our experiment, because our examples used different scenes.  
621 Similarly, we can reconcile the current results with research showing allocentric spatial direction  
622 decoding in RSC (Vass and Epstein, 2013; Marchette et al., 2015). RSC contains representations  
623 of (allocentric) spatial directions that are aligned with respect to a prominent direction in the  
624 environment (e.g., the major axis of a building (Marchette et al., 2015); or the direction of a  
625 distal landmark, like a city in the distance (Shine et al., 2016)). IPS, on the other hand, contains  
626 representations of (egocentric) spatial directions that are aligned with respect to an individual's  
627 current facing direction (Schindler & Bartels, 2013). Participants encoded directions in the  
628 current experiment egocentrically, supported by behavioral evidence – reaction time correlated  
629 with egocentric angle between the current and preceding spatial directions (but not visual angle).

630 We do not know if changing task instructions to promote allocentric direction coding (e.g.,  
631 asking participants to encode the spatial direction with respect to different sides of the screen),  
632 would yield cross-format spatial direction decoding in RSC.

633 Do schemas occupy a middle ground between words (abstract and arbitrarily related to  
634 the concept they denote), and visual scenes (concrete, rich in relevant and irrelevant detail)?  
635 Other conceptual domains support this notion of neural overlap between schemas, words, and  
636 visual depictions of concepts. Previous work on spatial prepositions report neural overlap in  
637 regions which process schemas and words, and separate areas which process schemas and visual  
638 images (Amorapanth et al., 2012). Viewing action words (like running) and schemas also  
639 resulted in cross-format decoding in action simulation and semantics areas (Quandt et al., 2017).  
640 In the current work, cross-format decoding of spatial directions was present in brain regions that  
641 process egocentric spatial directions.

642 Despite format being irrelevant for the task, format decoding was robust. Whereas images  
643 were processed distinctly from schemas and words in visual scene regions, schemas and words  
644 were disambiguated in IPS, as well as in object and visual word form areas. This pattern of  
645 results supports a model of concept coding in which abstract features are extracted from stimuli  
646 in format-dependent regions, then conveyed to brain regions which perform computations on the  
647 abstract concept. This finding is consistent with our behavioral data, suggesting implicit neural  
648 differences in the way scenes, schemas, and words are processed.

649 One limitation of our results is that we cannot account for all task-based effects (Harel et  
650 al., 2014) such as requiring that spatial directions be grouped into seven categories or that  
651 participants must respond to any of the three representational formats. We used a naturalistic task  
652 because of its applied relevance. When reading directions, for example, one might need to match

653 a 'slight left' from memory to an egocentric road direction, a task which is comparable to our  
654 rolling one-back design, and requires a navigator to translate words to scenes. Still, spatial  
655 directions are not always categorized discretely. During walking a human navigator can easily  
656 turn 145° clockwise, while not necessarily categorizing this turn as “sharp right.” Nevertheless,  
657 we observed spatially-specific categorization in bilateral IPS for visual scenes: lefts were more  
658 similar to each other than rights, excluding the exact same direction. Note that such processing is  
659 counter-productive for the one-back task. Representing slight left as more similar to left than to  
660 slight right means it is harder to disambiguate a slight left from a left.

661       In sum, the current experiments reveal similarities and differences in formats of spatial  
662 direction depictions. Behaviorally, people responded to schemas and words more quickly than  
663 visual scenes. Neural decoding of spatial directions for visual scenes occurred in IPS bilaterally.  
664 This region revealed evidence of cross-format, abstract representation of spatial directions. These  
665 data challenge the specificity of IPS in encoding egocentric spatial directions, and support a  
666 model of spatial processing wherein images involve spatial direction computation, whereas  
667 schemas and words do not.

## 668 References

669

670 Aguirre GK (2007) Continuous carry-over designs for fMRI. *Neuroimage* 35:1480–1494.

671 Amorapanth P, Kranjec A, Bromberger B, Lehet M, Widick P, Woods AJ, Kimberg DY,

672 Chatterjee A (2012) Language, perception, and the schematic representation of spatial

673 relations. *Brain Lang* 120:226–236.

674 Bonner MF, Epstein RA (2017) Coding of navigational affordances in the human visual system.

675 *Proc Natl Acad Sci U S A* 114:4793–4798.

676 Chen G, King JA, Burgess N, O’Keefe J (2013) How vision and movement combine in the

677 hippocampal place code. *Proc Natl Acad Sci* 110:378–383.

678 Epstein RA (2008) Parahippocampal and retrosplenial contributions to human spatial navigation.

679 *Trends Cogn Sci* 12:388–396.

680 Epstein R, Graham KS, Downing PE (2003) Viewpoint-specific scene representations in human

681 parahippocampal cortex. *Neuron* 37:865–876.

682 Etienne AS, Maurer R, Séguinot V (1996) Path integration in mammals and its interaction with

683 visual landmarks. *J Exp Biol* 199:201–209.

684 Galati G, Pelle G, Berthoz A, Committeri G (2010) Multiple reference frames used by the human

685 brain for spatial perception and memory. *Exp Brain Res* 206:109–120.686 Gilboa A, Marlatte H (2017) Neurobiology of Schemas and Schema-Mediated Memory. *Trends*687 *Cogn Sci* 21:618–631.

688 Harel A, Kravitz DJ, Baker CI (2014) Task context impacts visual object processing

689 differentially across the cortex. *Proc Natl Acad Sci* 111:E962–E971.

690 Haxby J V., Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and

691 overlapping representations of faces and objects in ventral temporal cortex. *Science*

- 692 293:2425–2430.
- 693 Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved optimization for the robust and  
694 accurate linear registration and motion correction of brain images. *Neuroimage* 17:825–841.
- 695 Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM (2012) FSL. *Neuroimage*  
696 62:782–790.
- 697 Julian JB, Fedorenko E, Webster J, Kanwisher N (2012) An algorithmic method for functionally  
698 defining regions of interest in the ventral visual pathway. *Neuroimage* 60:2357–2364.
- 699 Julian JB, Ryan J, Hamilton RH, Epstein RA (2016) The Occipital Place Area Is Causally  
700 Involved in Representing Environmental Boundaries during Navigation. *Curr Biol* 26:1104–  
701 1109.
- 702 Karnath HO (1997) Spatial orientation and the representation of space with parietal lobe lesions.  
703 *Philos Trans R Soc Lond B Biol Sci* 352:1411–1419.
- 704 Kirby JR, Moore PJ, Schofield NJ (1988) Verbal and visual learning styles. *Contemp Educ*  
705 *Psychol* 13:169–184.
- 706 Klippel A, Montello DR (2007) Linguistic and Nonlinguistic Turn Direction Concepts. In:  
707 *Spatial Information Theory*, pp 354–372. Berlin, Heidelberg: Springer Berlin Heidelberg.
- 708 Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping.  
709 *Proc Natl Acad Sci U S A* 103:3863–3868.
- 710 Marchette SA, Vass LK, Ryan J, Epstein RA (2015) Outside Looking In: Landmark  
711 Generalization in the Human Navigational System. *J Neurosci* 35:14896–14908.
- 712 Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: a  
713 primer with examples. *Hum Brain Mapp* 15:1–25.
- 714 Nili H, Wingfield C, Walther A, Su L, Marslen-Wilson W, Kriegeskorte N (2014) A Toolbox for

- 715        Representational Similarity Analysis. *PLoS Comput Biol* 10.
- 716    Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory.
- 717        *Neuropsychologia* 9:97–113.
- 718    Oliva A, Torralba A (2006) Building the gist of a scene: the role of global image features in
- 719        recognition. *Prog Brain Res* 155:23–36.
- 720    Poucet B (1993) Spatial cognitive maps in animals: new hypotheses on their structure and neural
- 721        mechanisms. *Psychol Rev* 100:163–182.
- 722    Quandt LC, Lee Y-S, Chatterjee A (2017) Neural bases of action abstraction. *Biol Psychol*
- 723        129:314–323.
- 724    Schindler A, Bartels A (2013) Parietal Cortex Codes for Egocentric Space beyond the Field of
- 725        View. *Curr Biol* 23:177–182.
- 726    Shine JP, Valdés-Herrera JP, Hegarty M, Wolbers T (2016) The Human Retrosplenial Cortex
- 727        and Thalamus Code Head Direction in a Global Reference Frame. *J Neurosci* 36:6371–
- 728        6381.
- 729    Talmy L (2000) *Toward a cognitive semantics*. MIT Press.
- 730    Vass LK, Epstein RA (2013) Abstract Representations of Location and Facing Direction in the
- 731        Human Brain. *J Neurosci* 33:6133–6142.
- 732    Vass LK, Epstein RA (2017) Common Neural Representations for Visually Guided
- 733        Reorientation and Spatial Imagery. *Cereb Cortex* 27:1457–1471.
- 734    Wang L, Mruczek REB, Arcaro MJ, Kastner S (2015) Probabilistic Maps of Visual Topography
- 735        in Human Cortex. *Cereb Cortex* 25:3911–3931.
- 736    Watson CE, Cardillo ER, Bromberger B, Chatterjee A (2014) The specificity of action
- 737        knowledge in sensory and motor systems. *Front Psychol* 5:494.

- 738 Whitlock JR, Sutherland RJ, Witter MP, Moser M-B, Moser EI (2008) Navigating from  
739 hippocampus to parietal cortex. *Proc Natl Acad Sci U S A* 105:14755–14762.
- 740 Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE (2014) Permutation inference  
741 for the general linear model. *Neuroimage* 92:381–397.
- 742
- 743

744 Legends.

745

746

**Figure 1. Stimuli and fMRI study paradigm.** Sample stimuli from the behavioral and

747 fMRI studies (A). Stimuli with phase shifted backgrounds are shown. A segment of the

748 experimental paradigm shown to participants in the fMRI study (B). In the fMRI study,

749 participants saw a spatial direction in one of the formats for 1 s, followed by a fixation cross for

750 2 s. During catch trials, the fixation cross instead disappeared after 500 ms, and a new spatial

751 direction appeared, in any of the formats, with the word "Same" underneath in red letters.

752 Participants pressed one button to indicate that the spatial direction on screen matched the one

753 previously seen, and the other button to indicate that it did not match (buttons were counter-

754 balanced across participants).

755 **Figure 2. Regions of Interest.** The five regions of interest (ROIs) used in the fMRI

756 analyses, shown for illustrative purposes (data analysis proceeded as described in the Methods).

757 ROIs are also displayed in the left hemisphere only, for ease of viewing, but were analyzed

758 bilaterally. The scene regions (Parahippocampal Place Area, PPA, yellow; Occipital Place Area,

759 OPA, green; Retrosplenial Cortex, magenta) are displayed as the top 100 voxels of the group

760 averaged t-statistic for the contrast of images - mean(words + schemas), and constrained by the

761 anatomical parcels from Julian et al. (2012). Intraparietal sulcus (IPS, red) and early visual

762 cortex (EVC, blue) were defined anatomically by the parcels from Wang et al. (2014). All ROIs

763 are displayed on the standard MNI map.

764 **Figure 3. Results from the behavioral study.** Response times were fastest overall for

765 schemas and words. Schemas also showed the largest within-format effect. That is, participants

766 were faster to respond when a schema came after a schema compared to word-word or image-

767 image.

768 **Figure 4. Within-format decoding of spatial direction.** The theoretical RDM (A) was  
769 compared to the neural RDM from five ROIs: the intraparietal sulcus (IPS), early visual cortex  
770 (EVC), and visual scene regions: occipital place area (OPA), parahippocampal place area (PPA),  
771 and retrosplenial complex (RSC). The IPS decoded spatial directions within images significantly  
772 greater than EVC. Within-format decoding was not significant in IPS or EVC for either schemas  
773 or words. Visual scene regions did not decode spatial direction within any of the three formats.

774 **Figure 5. Cross-format decoding of spatial direction.** The theoretical RDM (A) was  
775 compared to the neural RDM from five ROIs: the intraparietal sulcus (IPS), early visual cortex  
776 (EVC), and visual scene regions: occipital place area (OPA), parahippocampal place area (PPA),  
777 and retrosplenial complex (RSC). The IPS decoded spatial directions across all three formats, but  
778 only marginally greater than EVC. Cross-format decoding was not significant in IPS or EVC for  
779 either schemas or words. Visual scene regions did not decode spatial direction across formats.

780 **Figure 6. Decoding of format in ROIs.** The theoretical RDM (A) was compared to the  
781 neural RDM from five ROIs: the intraparietal sulcus (IPS), early visual cortex (EVC), and visual  
782 scene regions - occipital place area (OPA), parahippocampal place area (PPA), and retrosplenial  
783 complex (RSC). All regions significantly decoded the format of the representation, except for  
784 RSC. Multidimensional scaling plots (C) reveal that IPS separates all three formats whereas OPA  
785 distinguishes images from the other two. Arrows depict that categorical spatial directions (right,  
786 slight right, and sharp right collapsed as right arrows; left, slight left, and sharp left collapsed as  
787 left arrows; ahead as an up arrow).

788 **Figure 7. Decoding of format, whole brain searchlight.** The theoretical RDM from  
789 Figure 5A generated the contrast between same format minus different format correlations for  
790 images-images minus images-words/schemas (in hot colors) and for schemas-schemas and

791 words-words minus schemas-words (in cool colors). Image correlations were strongest in scene  
792 regions (OPA, PPA, RSC) and IPS, whereas schema and word correlations were strongest in  
793 word and object visual areas. Lower bound for searchlights are permutation corrected thresholds;  
794 upper bounds are  $p < .00001$  uncorrected.

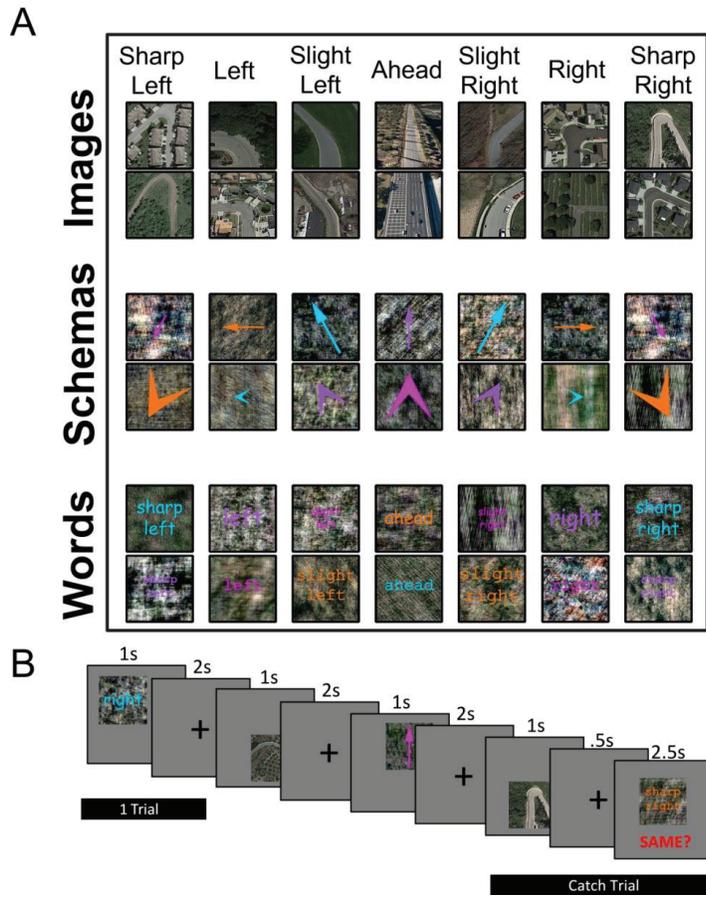
795 **Table 1.** IPS (Intraparietal Sulcus). EVC (Early Visual Cortex). OPA (Occipital Place  
796 Area). PPA (Parahippocampal Place Area).

797

798

799

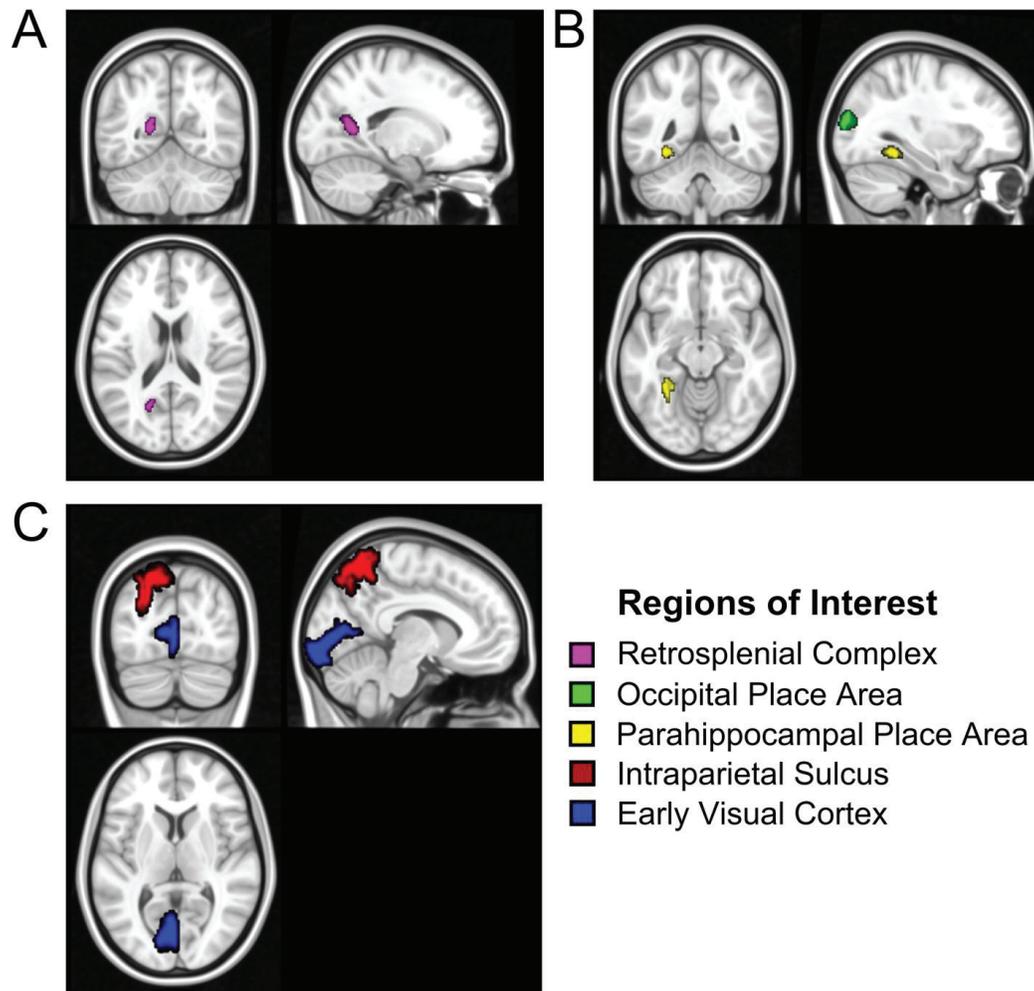
800



801

802

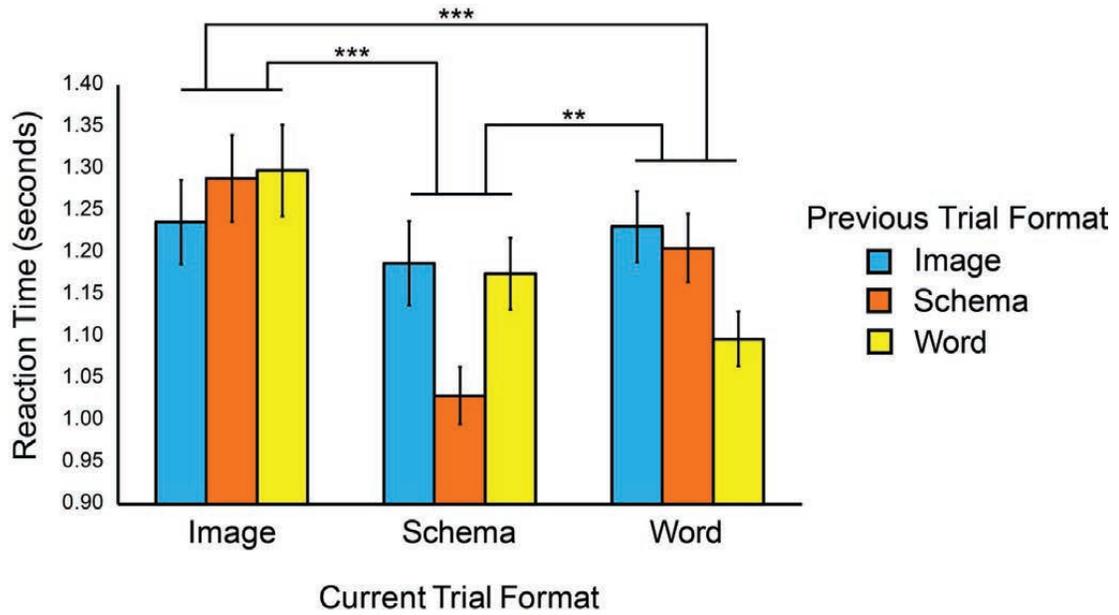
Figure 1.



803  
804  
805  
806  
807

**Figure 2.**

808



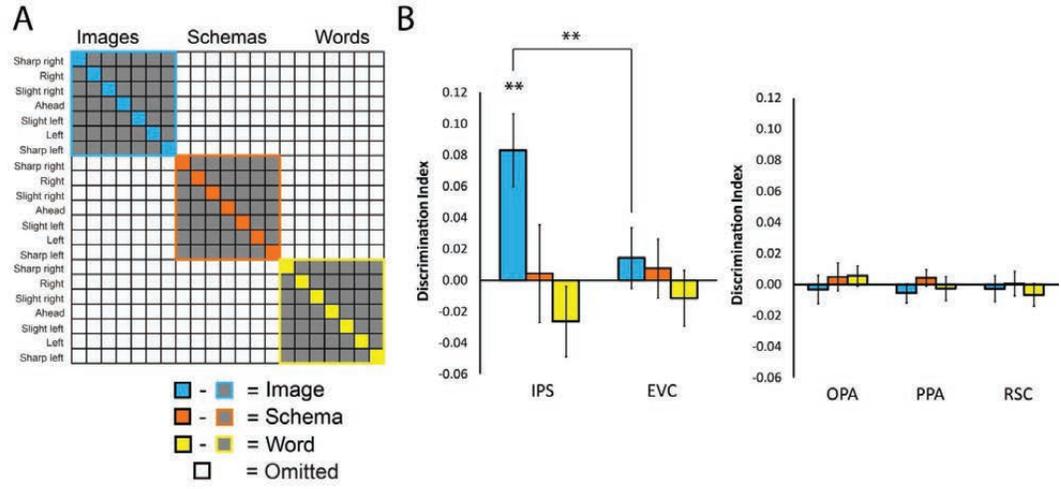
809

810

**Figure 3.**

811

812



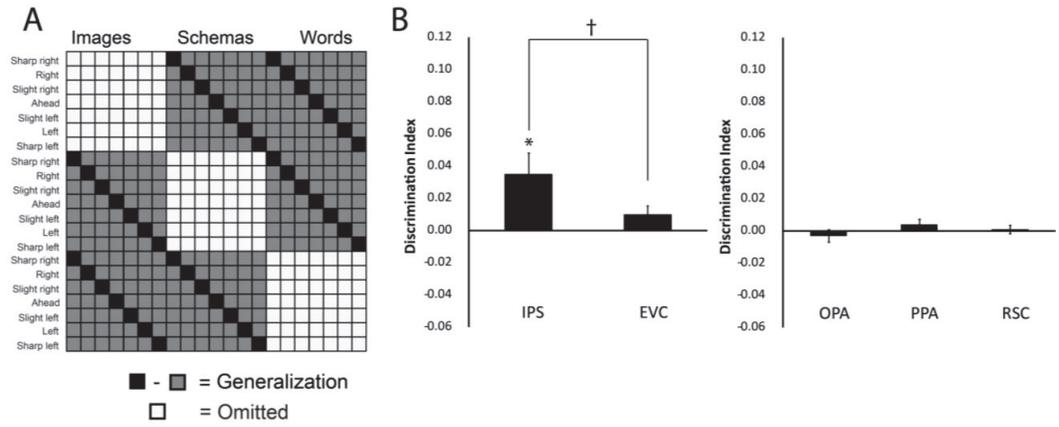
813

814

**Figure 4.**

815

816



817  
818

819 **Figure 5.**

820

821

822

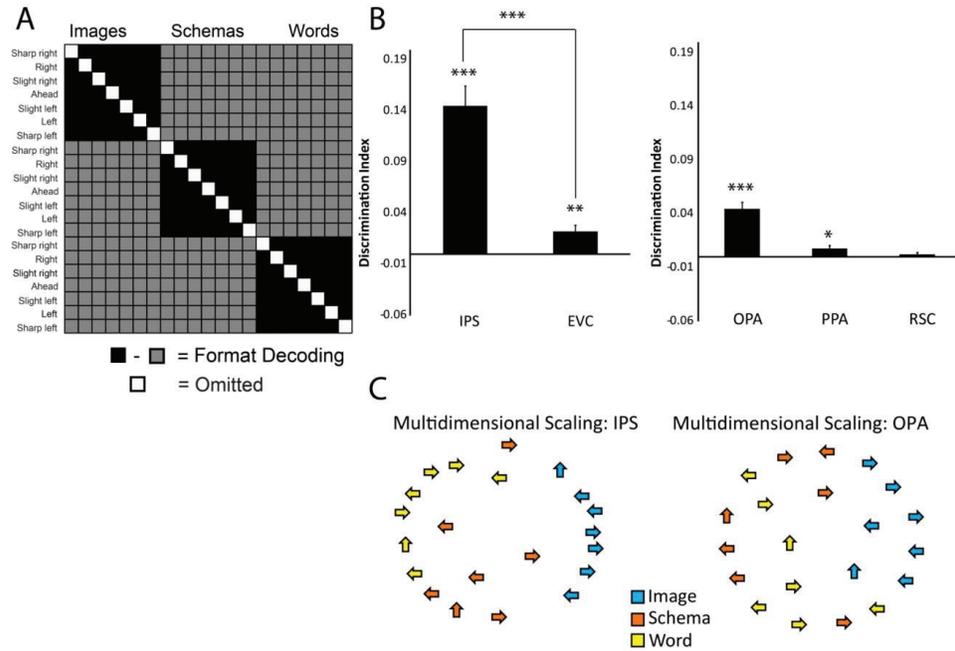
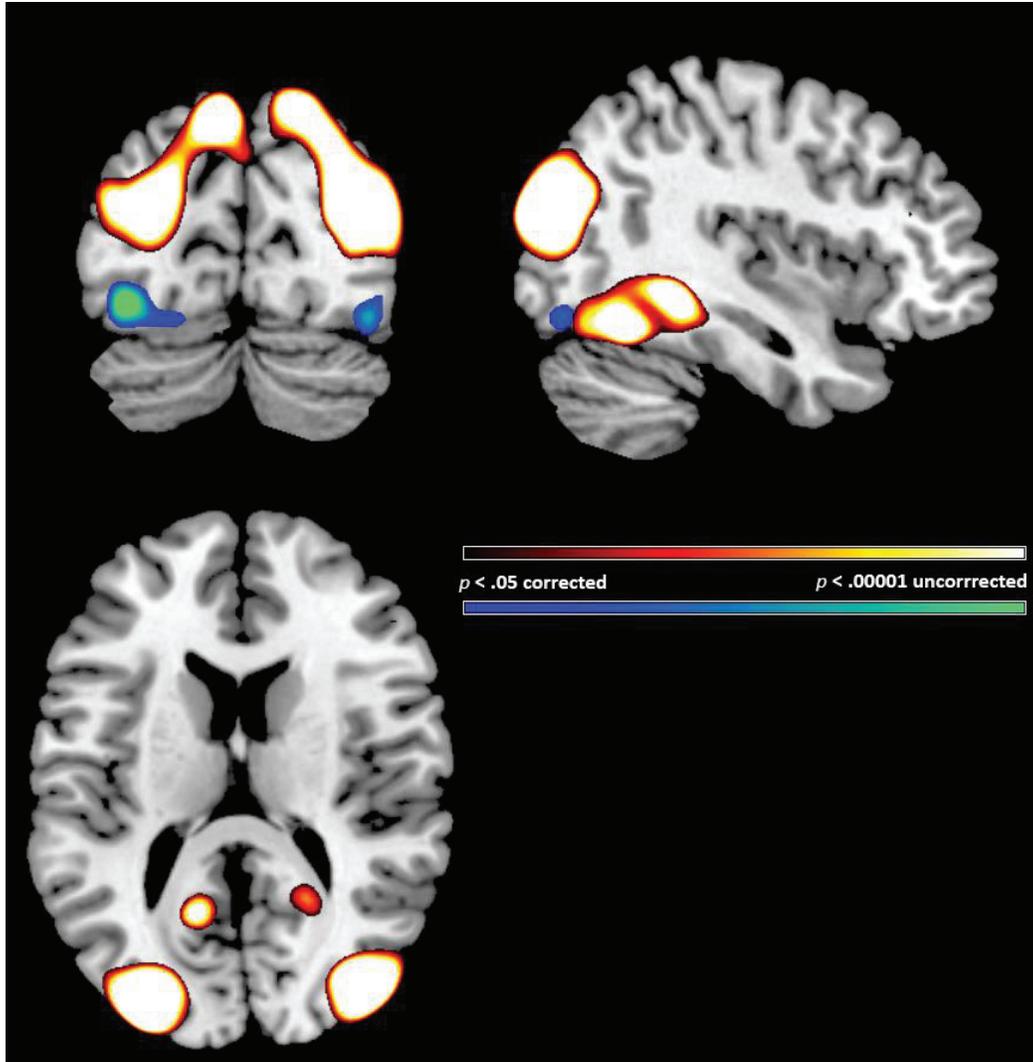


Figure 6.

823  
824

825

826

827  
828  
829  
830**Figure 7.**

831 **Table 1.** Pairwise format similarity in IPS, EVC, OPA, and PPA.

832

Brain region	Schema-Word		
	Mean (Standard deviation)	<i>p</i> -value	Effect Size ( <i>d</i> )
IPS	.034(.037)	0.0007	0.92
EVC	.025(.035)	0.0057	0.71
OPA	.006(.035)	0.017	0.17
PPA	-.0006(.009)	0.77	0.07
Schema-Image			
	Mean (Standard deviation)	<i>p</i> -value	Effect Size ( <i>d</i> )
IPS	.19(.127)	0.0000008	1.5
EVC	.027(.039)	0.0057	0.69
OPA	.06(.04)	0.000002	1.5
PPA	.013(.024)	0.026	0.542
Image-Word			
	Mean (Standard deviation)	<i>p</i> -value	Effect Size ( <i>d</i> )
IPS	.19(.17)	0.00003	1.12
EVC	.015(.0035)	0.061	0.43
OPA	.068(.049)	0.000005	1.39
PPA	.011(.020)	0.019	0.55

833

834



