

# Decoding individual identity from brain activity elicited in imagining common experiences

Andrew James Anderson <sup>1,2</sup>✉, Kelsey McDermott<sup>3,4</sup>, Brian Rooks<sup>3</sup>, Kathi L. Heffner<sup>3,5,6</sup>, David Dodell-Feder <sup>1,2,7</sup> & Feng V. Lin<sup>1,2,3,8,9</sup>

Everyone experiences common events differently. This leads to personal memories that presumably provide neural signatures of individual identity when events are reimagined. We present initial evidence that these signatures can be read from brain activity. To do this, we progress beyond previous work that has deployed generic group-level computational semantic models to distinguish between neural representations of different events, but not revealed interpersonal differences in event representations. We scanned 26 participants' brain activity using functional Magnetic Resonance Imaging as they vividly imagined themselves personally experiencing 20 common scenarios (e.g., dancing, shopping, wedding). Rather than adopting a one-size-fits-all approach to generically model scenarios, we constructed personal models from participants' verbal descriptions and self-ratings of sensory/motor/cognitive/spatiotemporal and emotional characteristics of the imagined experiences. We demonstrate that participants' neural representations are better predicted by their own models than other peoples'. This showcases how neuroimaging and personalized models can quantify individual-differences in imagined experiences.

<sup>1</sup>Department of Neuroscience, University of Rochester Medical Center, Rochester, NY 14642, USA. <sup>2</sup>Del Monte Institute for Neuroscience, University of Rochester Medical Center, Rochester, NY 14642, USA. <sup>3</sup>Elaine C. Hubbard Center for Nursing Research on Aging, School of Nursing, University of Rochester Medical Center, Rochester, NY 14642, USA. <sup>4</sup>Neuroscience, University of Arizona, Tucson, AZ 85721, USA. <sup>5</sup>Department of Psychiatry, University of Rochester Medical Center, Rochester, NY 14642, USA. <sup>6</sup>Division of Geriatrics and Aging, Department of Medicine, University of Rochester Medical Center, Rochester, NY 14642, USA. <sup>7</sup>Department of Psychology, University of Rochester, Rochester, NY 14642, USA. <sup>8</sup>Department of Neurology, University of Rochester Medical Center, Rochester, NY 14642, USA. <sup>9</sup>Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14642, USA. ✉email: [aander41@ur.rochester.edu](mailto:aander41@ur.rochester.edu)

Almost everyone can imagine themselves at a wedding, however each person does so differently because they have been to different weddings and hence draw upon memories that no one else has. Our personal history of episodic memories contributes to defining us as individuals and in extreme cases—where memories are of traumatic events—can profoundly affect our psychological health and quality of life. A principal goal of cognitive science is to understand how such memories are represented and manipulated in the human brain. Memories of past experiences can be activated and relived through recollection and are thought to be pieced together to support the mental simulation of hypothetical scenarios<sup>1–5</sup>. Functional brain imaging now enables the systematic study of brain activation elicited during recall and imagination. A long-term vision for the future might therefore be to devise technology that provides a comprehensive neural readout of the information that one voluntarily activates in mental imagery. More humble prerequisites for this are to establish that neural activity elicited during imagery captures meaningful differences between different individuals' mental simulations of similar kinds of events, and to devise quantitative methods that can predict the information represented in neural activation.

In previous work, functional Magnetic Resonance Imaging (fMRI) studies of brain activity have identified a core distributed network of neuroanatomical regions that are reliably activated during the recollection and/or imagination of different experiences and scenarios<sup>6–8</sup>. Regions that are activated in episodic recollection and simulation strongly overlap, which suggests that similar neural machinery is engaged in both cases<sup>1,3–5</sup> (although activation patterns elicited in remembering, and imaging possible past and future events are still distinguishable<sup>5</sup>). This core episodic recollection and simulation network includes regions of medial parietal cortex, inferior parietal cortex, medial prefrontal cortex, and medial and lateral temporal lobe<sup>1–13</sup>. Researchers seeking to decipher what information is represented in brain activity within regions of this network have shown that different types of event can be distinguished from multiple network regions<sup>14–17</sup>, as well different components of individual events<sup>18–24</sup> (e.g., people, places, objects, space/time of occurrence). However, it has remained unclear whether differences in activation patterns between individuals imagining similar types of event represent anything more than functionally-irrelevant between-subject noise.

Reason to hypothesize that neural correlates of person-specific imagery may be readable from fMRI data comes in part from a recent thrust of research that has exposed individual differences in other cognitive domains<sup>25</sup>. Prominent examples include: using resting-state fMRI data to predict individual differences in the cortical distribution of activity elicited in multiple (gambling/motor/language etc.) task-related studies<sup>26</sup>, matching representational similarity of fMRI elicited viewing photos of personal belongings to behavioral measures of object similarity<sup>27</sup>; and using measures of brain activity scanned during a picture-based mechanical engineering task to predict individuals' physics/engineering exam results<sup>28</sup>. Research has even begun to characterize brain network activity in terms of the nature of individual differences in ongoing thought, reflecting whether current thoughts are detailed, correspond to the past or future, are verbal or in images and so on<sup>29,30</sup>. However, whilst these studies do demonstrate that individual differences can be discerned using fMRI, it is unclear that the methods utilized to elicit individual-differences (e.g., picture interpretation, resting) and detect them (e.g., exam results, object similarity judgments, ratings on components of thought) would generalize to the current case of imagining oneself in multiple different scenarios when cued by generic prompts such as “a wedding” or “a funeral”.

In addition, a number of studies have cataloged between-group differences in autobiographical memory-related fMRI activation. For instance, altered activation has been observed in Alzheimer's disease<sup>31,32</sup>, semantic dementia<sup>33</sup>, and epilepsy<sup>34</sup>. Relating more to conceptual knowledge, machine learning methods have discriminated between young suicidal ideators thinking about suicide and emotion concepts and healthy controls<sup>35</sup>, and students contemplating objects' mechanical function at different phases of education<sup>36</sup>. However, whilst all of the previous between-group differences must be built on top of an accumulation of individual-differences (neural features that appear in one group but not the other), and whilst both group-average and trait-level individual difference studies have revealed the engagement of similar brain networks<sup>30,37,38</sup>, group-averages cannot explain detailed differences in event representations between pairs of individuals from within the same group.

A question is then posed over how personal signatures of imagined experiences can be identified in fMRI data, and how the information content of brain activity can be interpreted. This would appear to rely on having personalized models of mental imagery that are sufficiently detailed to not only capture differences between events, but also between individuals' idiosyncratic experiences of similar events. Modeling advances<sup>39–45</sup> have led to the ability to predict/decode fMRI activity elicited as third-party sentence-level descriptions of events are read/heard<sup>46–53</sup>. However, such neural decoding models have largely relied upon internet-based approaches to model generic representations of meaning by computationally harvesting semantic representations from massive text repositories, or crowd-sourcing behavioral ratings across large collectives of people. Consequently, the resultant models are built at group-level, and whilst they have high signal-to-noise-ratio when it comes to capturing broad population-level commonalities, they have zero ability to discern individual differences. It is unknown whether current modeling approaches can be tailored to be practically effective at capturing interpersonal differences in imagined experiences. It is also unknown how accurately these current modeling approaches can explain neural activity elicited as individuals vividly imagine situations from their own perspective, as opposed to passively reading third-party sentences<sup>49–53</sup>.

We here hypothesize that fMRI activity patterns elicited as people imagine weddings, funerals and eighteen other common scenarios reflect interpersonal differences in the information that is brought to mind. We test this by devising personal models of the imagined experiences to discern participant identity from corresponding fMRI data. In so doing, we provide evidence that: (1) Neuroimaging can quantitatively measure meaningful individual differences in brain activity elicited as people imagine complex events from their own perspective. (2) The information content of brain activation can be predicted using personalized models derived from verbal descriptions and behavioral ratings of the imagined events. We discuss the potential implications this has for both basic and applied science.

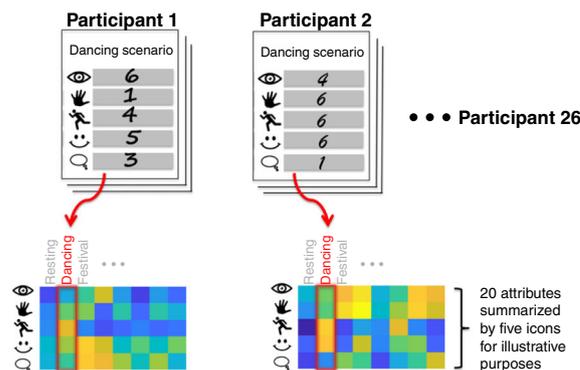
## Results

We selected 20 common scenarios (e.g., reading, dancing, restaurant) that we asserted would be common experiences to our study population, whilst also being of a sufficiently broad nature that different people would have different experiences of each (e.g., most people have experienced dancing, whether it was actively or passively, in company or isolation, or enjoyable). We next recruited 30 individuals who were healthy agers, although 26 completed the experiment (17F, 9M, mean  $\pm$  SD age = 73  $\pm$  7years). Healthy agers were chosen in particular to evaluate the feasibility of the approach for future tests on clinical aging

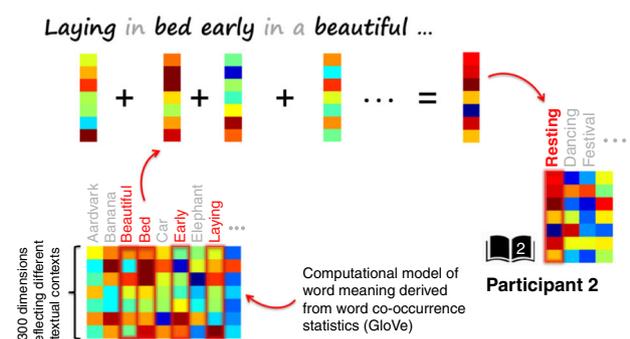
1. 26 participants vividly imagined and then verbally described 20 common scenarios



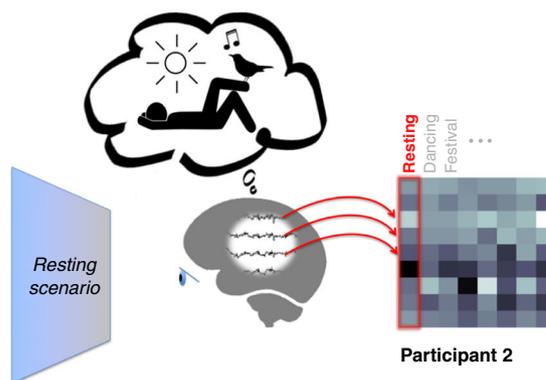
2. Participants individually rated their imagined scenarios on 20 experiential attributes



3. Verbal descriptions were mapped to a text-based distributional semantic model of word meaning



4. Participants underwent fMRI as they reimagined the scenarios when prompted by standardized cues



**Fig. 1 Data collection: experimental protocol and construction of personal models of the imagined scenarios.** This diagram summarizes the entire data collection procedure as 4 stages, ranging from the collection and processing of behavioral data to fMRI scanning. The chronological order of stages 1, 2, and 4 reflects the actual order of experimentation. Stage 3 is automated and could in practice take place at any time after stage 1.

populations. In the Discussion we consider what similarities and differences might arise from a comparative study on younger adults, although this remains an open empirical question for the time being.

Each participant attended a visit with an experimenter, who read the different scenarios to the participant one by one (e.g., “A dancing scenario”). On each reading the participant was requested “to vividly imagine themselves in the scenario” placing importance on their (imaginary) perception, action and feelings (Fig. 1). Participants then provided a verbal sentence length description of each imagined scenario. Examples illustrated in Fig. 1 indicate that the descriptions supplied by different participants for the same scenario were at least superficially different. To estimate the extent to which participants could base imagination on real past experiences of the event in question (as opposed to envisioning entirely new fictitious situations) we asked participants to rate this, as well as how vividly they could imagine each scenario. Participants tended to rate scenarios to have happened and to have been vividly imagined (grand mean  $\pm$  SD ratings across both participants and scenarios were  $5.5 \pm 1.1$ , and  $5.2 \pm 1.2$ , respectively, on a Likert scale of [0 6]). See Supplementary Information for rating statistics for individual scenarios (Supplementary Fig. 1) and details of the ratings procedure.

**Overview of personalized models of imagined experiences.** To model the person-specific information content of each

participant’s imagined scenarios, we employed a joint-modeling approach that integrated semantic information that was estimated from participants verbal descriptions of events with non-linguistic sensory, motor, interoceptive, emotional, locational and temporal information that could putatively contribute to episodic recall and simulation (as there is evidence for at least in conceptual representation<sup>53–56</sup>). The linguistic and non-linguistic models were both founded on established approaches that have been extensively used in modeling the semantic representation of words and concepts since refs. 39–45. Different to previous work that has constructed these models at group-level, we here newly adapted the modeling approaches to capture person-specific scenario representations and newly leveraged them to explain (person-specific) neural activation elicited as participants actively imagined themselves in common scenarios. See also ref. 57 for a language modeling approach to representing individuals and ref. 21 for a personal image-tagging approach to associating autobiographical bodycam photographs with experience).

The linguistic modeling approach, henceforth referred to as the “verbal model”, borrowed a popular method from computational linguistics to transform participant’s verbal scenario descriptions into quantitative representations reflecting the linguistic meaning of each description. Specifically, we took a state-of-the-art distributional semantic model (GloVe<sup>44</sup>) that approximates words’ meaning in terms of the textual contexts that words appear in as measured across a large text corpus. For instance, pyramids and camels are related because they co-occur together

within sentences, and camels and donkeys are related because they occur independently in “riding” sentences. Practically each word is represented as 300-dimensional vector, derived by factorizing a word co-occurrence matrix (vocabulary size is 2.2million words and co-occurrences were measured across 840 billion tokens from Common Crawl <https://commoncrawl.org>). We cut out content words (words with intrinsic meaning as opposed to a grammatical function) from participants’ verbal descriptions, and mapped each content word to the corresponding GloVe vector. We then combined word-vectors together to model entire scenario descriptions through pointwise addition. Although such additive composition is a naïve strategy that discounts word order and syntax, it has proven practically effective in both computational linguistics<sup>58,59</sup> and in predicting sentence-level brain activation<sup>49–53</sup>. This process yielded a personalized distributional semantic model representation of each participant’s verbal descriptions of each scenario for each participant (even though word vectors were drawn from GloVe in each case).

A limitation of verbal descriptions is that they may omit information that is so salient in physical experience that it is assumed to be too obvious to bother expressing. To cite a common example, people seldom feel the need to communicate the color of bananas<sup>60</sup>, or that they played tennis in daylight, or watched a movie in the dark, because this is assumed to be common knowledge. To counteract this limitation and potentially build a more comprehensive model of episodic imagery, we estimated the degree that different sensory/motor/cognitive/emotional neural systems would have been engaged if the participant had physically experienced the scenario, and hence might have contributed information to memories and in turn episodic simulation. To this end we constructed an experiential attribute model<sup>45</sup>. Specifically, the attribute model was implemented by having each participant rate each scenario on 20 sensory, motor, cognitive, spatiotemporal and emotional experiential attributes (Likert scale 0–6, see Fig. 1, see “Methods” section and Supplementary Information for specific details). For instance, a regular dancer might be expected to imagine themselves actively engaging at a dancing event, and hence rate it strongly on motor engagement and positive affect. In contrast, an ardent dance-avoider might imagine and rate the opposite. The set of 20 attributes was an abridged selection of a wider set of 65 attributes identified by ref. <sup>45</sup> that when crowd-sourced at group-level, had provided a basis for predicting brain activity elicited as event-related sentences were read<sup>49,52,53</sup>. The 65 attributes were cut down to 20 before experimentation to meet timing constraints (see “Methods” section for further details). Importantly, the attribute model explicitly solicited for a comprehensive rundown of experiential information that participants might not have thought to provide in their verbal reports if left to themselves (even if their reports were open ended in length).

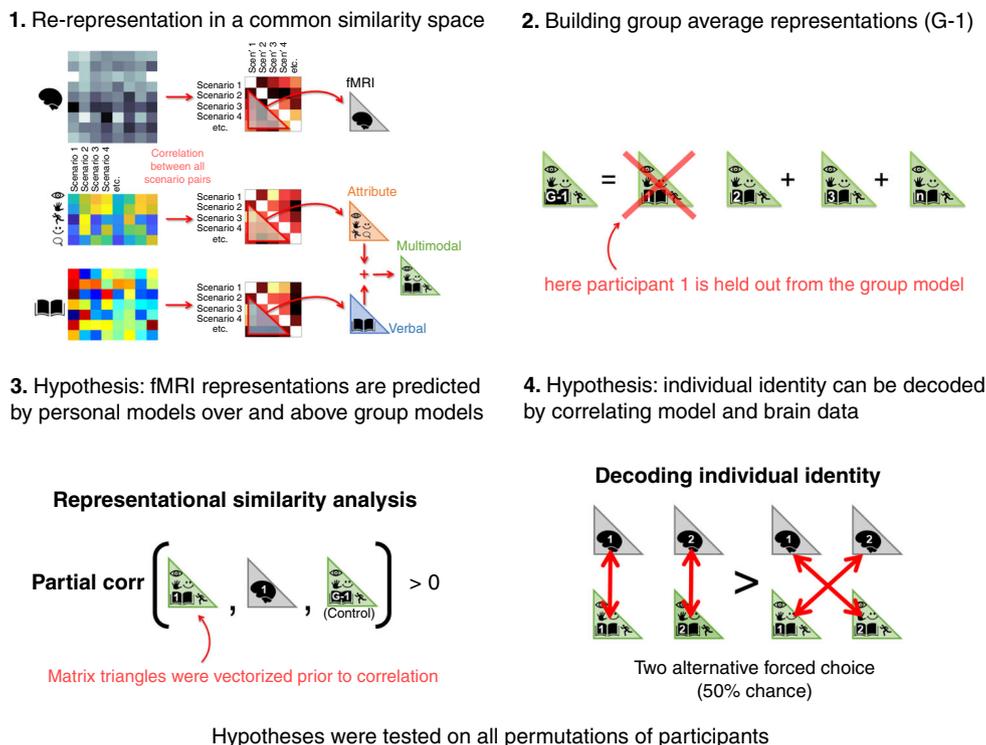
**fMRI experiment overview.** Participants’ brain activity was scanned using fMRI as they vividly re-imagined the scenarios (Fig. 1). During fMRI participants imagined their own scenario when a standard written prompt (“A dancing scenario”) was visible. To accommodate noisy fMRI measures, the set of 20 scenario prompts were displayed 5 times over (in different random orders), and participants reimagined the same scenarios each time. fMRI data was preprocessed using standard techniques to correct for slice timing and head motion and each participants’ fMRI data was spatially normalized to a common anatomical template (MNI). Previous work<sup>14–24,49,52,53</sup> led us to expect that activation patterns within localized brain regions would

comprehensively represent multiple scenarios (in particular the episodic recollection/simulation network<sup>6–8</sup>). Thus, we parcellated fMRI activation the cortex into 90 neuroanatomical regions of interest (ROIs) using Automated Anatomical Labeling<sup>61,62</sup>. We analyzed all regions (rather than a predefined network) to avoid the possibility of overlooking patterns of interest. We estimated the top 100 most informative voxels per ROI using a frequently used voxel stability criteria<sup>46</sup> (see “Methods” section). Comparative results using 50 and 200 voxels yielded a similar pattern of results, and are included in Supplementary Information. The 5 fMRI volumes per scenario were voxel-wise averaged together and vectorized to provide a single 100dimensional vector per scenario, per ROI, per participant.

**fMRI activation patterns reflect interpersonal differences in scenario representation.** To test our overarching hypothesis that fMRI activation patterns elicited in imagining the scenarios reflected interpersonal differences, we applied Representational Similarity Analysis (RSA)<sup>63</sup> to compare model and fMRI data (Fig. 2 and see “Methods” section). In the case at hand, RSA can be conceptualized as a two-step process that begins by re-representing each of the 20 scenarios as an individual point in a geometric similarity space. This similarity space is common across model and fMRI data to enable their comparison, but the coordinates of the scenarios are defined separately for the model and fMRI data. The second step tests whether the relative geometric arrangement of the scenario coordinates is similar for the model and fMRI data. For instance, if in fMRI similarity space it turns out that the dancing scenario is close to party and wedding, whilst reading is close to writing, and cooking is close to housework, and this same relationship is also seen in model similarity space, then there is evidence that this arrangement is meaningful because the model predicts the representational structure of the fMRI data. Practically, the transformation to similarity space was implemented by computing Pearson inter-scenario correlation matrices separately for model and fMRI data. Model to fMRI similarity space comparisons were implemented by computing Spearman correlation between vectors of values that define the respective fMRI/model correlation matrices (high correlation coefficients reflect a strong relationship, see “Methods” section for more details and a description of the statistical significance testing procedure).

The main interest of the current analysis, was however on detecting meaningful representational structure in fMRI data that is peculiar to each individual, and not a feature of the general population. For instance, if a participant (unusually) reported that they imagined themselves dancing whilst doing their housework then their corresponding model might predict an unusually high similarity between fMRI representations of housework and dancing. To this end, we computed group-average model representations to estimate population-level commonalities in scenario representational similarity structure, and used these as a control in our forthcoming analyses that sought to identify person-specific idiosyncrasies.

All of the primary analyses reported (Fig. 2) incorporated our three personalized data types: (1) fMRI activation (illustrated in gray with the brain icon). (2) The attribute model (illustrated in orange with icons for attributes). (3) The verbal model (illustrated in blue with a book icon). To create a single personal multimodal model representation (illustrated in green with attributes and book icons) we integrated information across the person-specific verbal and attribute models by averaging representations in similarity space (see Fig. 2 and “Methods” section for details, and Supplementary Information for evidence that model integration was beneficial). Group-average models were estimated to serve as



**Fig. 2 Representational similarity analysis protocol and hypothesis formalization.** The diagram summarizes the computational approach taken in our analyses and illustrates two formalizations of our primary hypothesis. Stage 1 was repeated for each participant. In the diagram fMRI data is represented as a single correlation matrix to simplify visualization, however stages 1 to 4 were repeated on fMRI activation extracted from multiple regions of interest (ROIs). Step 1 illustrates how the three personal scenario representations (fMRI and the verbal and attribute models) were transformed into a common space to enable their comparison (stages 3 and 4). Step 2 illustrates how group-average model representations were constructed whilst excluding test participants (G-1). Stages 3 and 4 reflect different approaches to testing the same fundamental hypotheses that fMRI activation patterns reflect meaningful person-specific information that is brought to mind when individuals imagine themselves in different scenarios.

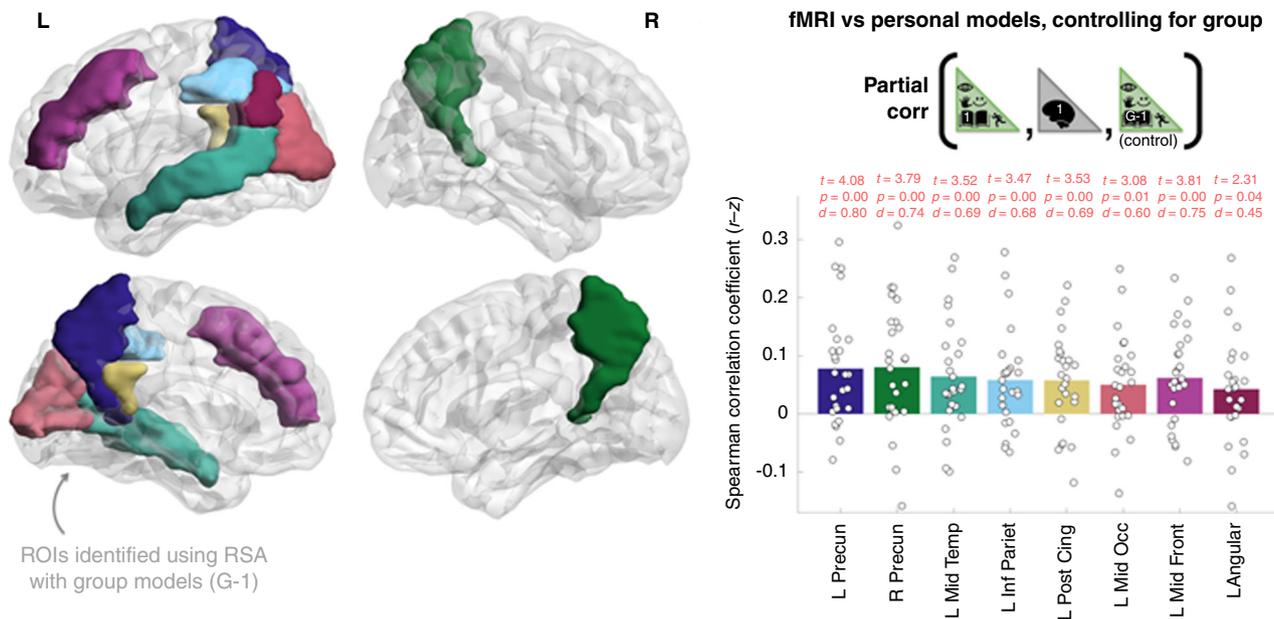
controls in our analyses in a similar fashion by averaging person-specific model representations in similarity space (see Fig. 2 and see “Methods” section for details). Critically a different group-average model was computed for each participant, that excluded the respective participant’s model data. So, in a test of participant 1’s fMRI, a group-level model (G-1) was built by averaging model similarities across participants 2 to 26 and so on. See “Methods” section for further details. Note, that group-average representations could equally have been built by averaging individuals in model feature space prior to computing similarity matrices, and this approach yields similar results (Supplementary Information).

We first questioned whether fMRI activation patterns in brain regions that reflect group-average model structure (i.e., what we would expect to predict using traditional group-level modeling approaches) also encode person-specific representations. To this end, we initially identified a network of eight brain regions that reflected group-average model structure (excluding the test participant) as is described in detail in the “Methods” section. A complete listing of RSA results for all ROIs is in Supplementary Table 1. The eight ROIs spanned left/right medial parietal cortex, left inferior parietal cortex, left lateral temporal cortex and left mid frontal cortex (the anatomical locations of these ROIs are illustrated in Fig. 3 left). We next tested whether participant-specific models could predict fMRI representational similarities (in the same participant’s brain) over and above the corresponding group-average model for the eight ROIs. This was implemented by computing the partial correlation (Spearman) between person-specific model similarity vectors and fMRI similarity vectors, whilst controlling for group-average model similarity vectors. Partial correlation coefficients were  $r$ -to- $z$  transformed. Then, for each of the eight ROIs, the set of

coefficients corresponding to the 26 participants were compared to zero using one sample  $t$ -tests (1-tailed, anticipating positive correlation). The resultant vector of eight  $p$ -values were corrected for multiple comparisons according to False Discovery Rate (FDR)<sup>64</sup>. Critically, all eight ROIs yielded FDR corrected  $p$ -values  $< 0.05$ . See Fig. 3 right for comprehensive test statistics and Supplementary Table 2 for a complete listing of results on all ROIs. These results provide evidence that person-specific elements of scenario representation were encoded in left/right medial parietal cortex, left inferior parietal cortex, left lateral temporal cortex and left mid frontal cortex. These regions broadly overlap with the core episodic recollection/simulation network, as considered further in the Discussion.

RSA coefficients computed using the personal models in isolation are illustrated in Fig. 4 left. Mean RSA coefficients for the personal models were marginally greater than the group-average models in 7/8 ROIs (Fig. 4, Supplementary Table 1) but there were no statistically significant differences for individual ROIs. The broad similarities in the magnitude of RSA coefficients between the personal and group modeling approaches probably reflect their complementary strengths. Specifically, whilst group models cannot capture the person-specific idiosyncrasies documented above, they have higher signal-to-noise for estimating population-level commonalities. We leave an explicit quantification of the complementary contribution made by the group-models over to future work because it is not the prime focus of the current article.

**Detailed neuroanatomical distribution of person-specific representational structure.** To follow up our previous ROI-based analysis we performed a searchlight analysis<sup>65</sup> to more precisely



**Fig. 3 fMRI activation patterns elicited in imagining common scenarios reflect person-specific information.** The plot shows how participant-specific scenario models predict the representational similarities of corresponding fMRI data over and above group-average models (that excluded the respective participant in each test). Black circles illustrate RSA coefficients for the 26 participants. Bar heights correspond to the mean across participants. Bar colors correspond to the ROIs illustrated on the brain (left). One sample t-tests tested whether partial RSA coefficients were greater than zero (1-tailed). Cohen's *d* was computed by dividing the t-statistic by  $26^{1/2}$ . Exact FDR<sup>64</sup> corrected p-values in the same order as plotted above were: 0.0030, 0.0030, 0.0034, 0.0034, 0.0034, 0.0077, 0.0030, 0.0400 (see also Supplementary Table 2). Permutation-based p-values derived from partial RSA tests performed at an individual-level revealed significant outcomes ( $p < 0.05$ ) in the following numbers of participants per ROI (in parentheses): L Precun (7), R Precun (9), L Mid Temp (5), L Inf Pariet (4), L Post Cing (4), L Mid Occ (3), L Mid Front (6), L Angular (4). The cumulative binomial probability of achieving 4 or more outcomes at  $p < 0.05$  in an ROI is 0.04. The eight ROIs presented were selected using G-1 data (see "Methods" section, the following figure and Supplementary Table 1 for results using the personal and group-average models separately). Brain illustrations were made using ref. <sup>103</sup>. Source data are provided as a Source Data file.

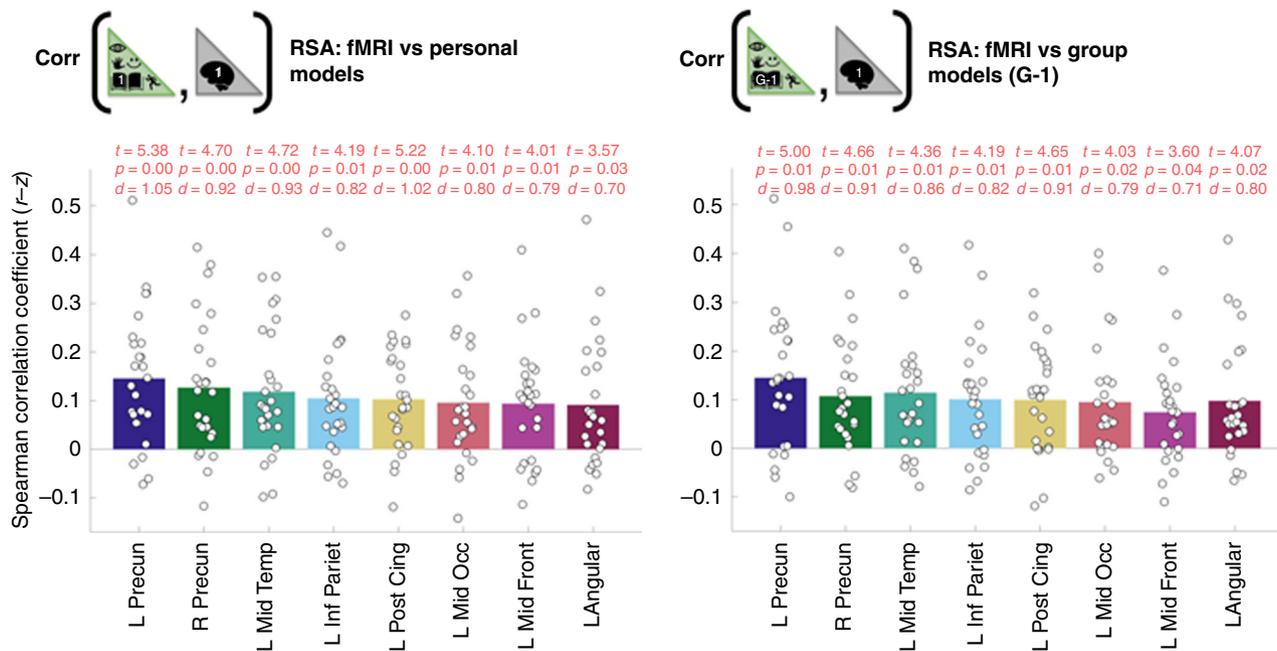
estimate the neuroanatomical layout of brain regions reflecting person-specific and group-average model structure. To this end we replicated the previous RSA and partial RSA within searchlight ROIs. Searchlight ROIs were cubes of radius 3 voxels (side 7) that were iteratively centered on every location in the brain via the implementation in ref. <sup>66</sup> (as is analogous to shining a searchlight, see "Methods" section). This complemented the previous anatomical ROI-based analysis which was well equipped to detect the presence of person-specific information in fMRI data (because informative voxels were selectively analyzed and weaker assumptions were placed on the shape of patterns of interest, beyond that they could fit into the relatively large anatomical ROIs). However, the previous ROI analysis did not precisely locate person-specific representations.

Results of the searchlight analyses are illustrated in Fig. 5, and the neuroanatomical locations of significant clusters ( $p < .05$  FDR<sup>64</sup> corrected) are identified in Supplementary Tables 3–5. Similar to the ROI-based analyses fMRI representations in medial parietal cortex and inferior parietal cortex were identified by the group-average models (Fig. 5 right) and were subsequently found to reflect person-specific information structure in the partial RSA analysis (Fig. 5 left). Different to the ROI-based analysis, prefrontal regions were not detected and inferior parietal cortex was less well represented, which may reflect a lower sensitivity of the searchlight approach (which did not exclude non-informative voxels, and may have been disadvantaged in capturing patterns that did not adequately fit within the searchlight). The searchlight RSA based on person-specific models alone (i.e., when the group-average was not controlled for) revealed a widely distributed network of brain regions that included clusters in dorsal and ventral medial prefrontal cortex, dorsolateral frontal cortex and

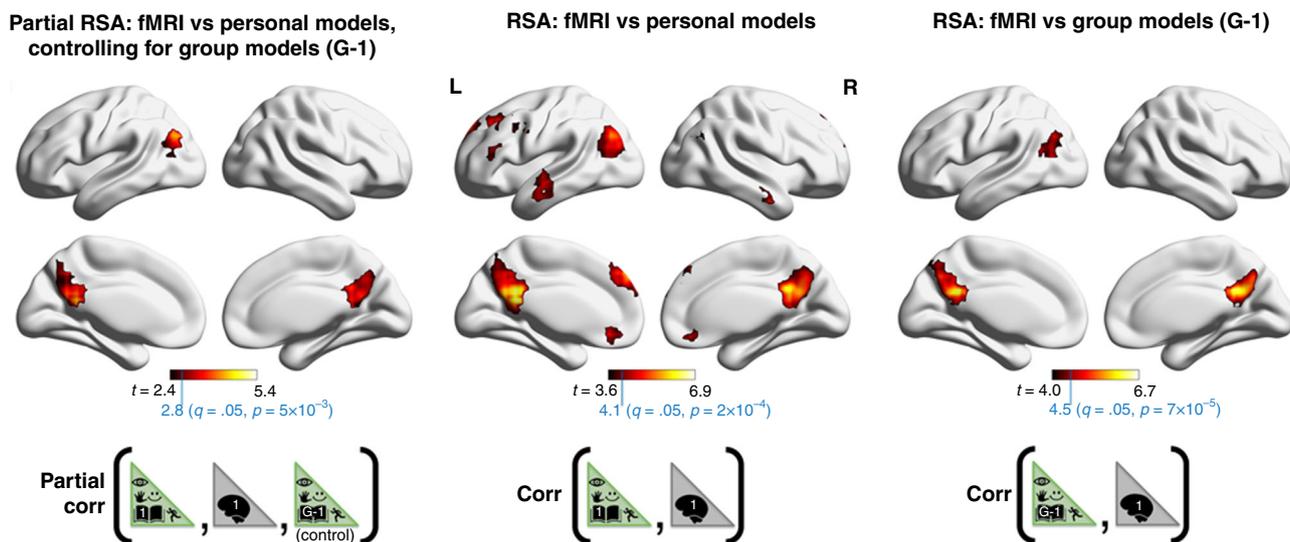
anterior temporal cortex (Fig. 5 middle). Importantly, this more neuroanatomically precise estimate of the distribution of regions encoding scenario information echoes the configuration of the core episodic simulation/recollection network<sup>6–8</sup> more precisely than the ROI-based analysis.

**Individual identity can be decoded from brain activity elicited imagining personal experiences.** As a second formalization of our key hypothesis, we explicitly tested how well individual identity could be decoded by matching participants' fMRI activation to their person-specific models. This analysis was repeated on each of the eight anatomical ROIs that had been identified in our earlier analyses. In advance, some degree of success in interpersonal decoding is already entailed from the earlier partial RSA analyses that revealed person-specific neural representations. However, the current analysis is still required to estimate just how accurately pairs of individuals can be distinguished, which would be difficult to estimate otherwise.

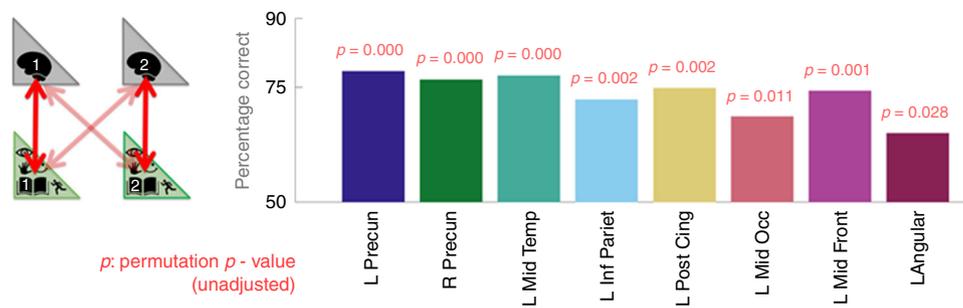
The decoding procedure is detailed in "Methods" section. In brief, pairs of test participants were drawn, and we computed RSA between each participant's person-specific model, and both participants' fMRI data (4 coefficients in total). We then tested whether RSA coefficients between participant-specific models and the corresponding fMRI data sets (P1-model vs. P1-fMRI + P2-model vs. P2-fMRI) were greater than the incongruent mismatch (e.g., P1-model vs. P2-fMRI + P2-model vs. P1-fMRI). A decoding accuracy was estimated as the percentage of times that congruent scores were greater than incongruent scores across all 325 participant pairs. If there were no participant-specific relationship an accuracy of 50% was expected.



**Fig. 4 RSA coefficients for personal models (capturing idiosyncratic features) and group-average models (high signal-to-noise for group-commonalities) were of broadly similar magnitudes.** Black circles illustrate RSA coefficients for the 26 participants. Bar heights correspond to mean values across participants. t-tests tested whether RSA coefficients were greater than zero. Cohen’s d was computed by dividing the t-statistic by  $26^{1/2}$ . P-values illustrated were FDR-corrected<sup>64</sup> across 90 ROIs. Exact FDR corrected p-values in the same order as plotted above for the personal models were: 0.0024, 0.0047, 0.0047, 0.0116, 0.0024, 0.0124, 0.0138, 0.0327. Exact FDR corrected p-values in the same order as plotted above for the group-average models were: 0.0070, 0.0070, 0.0111, 0.0140, 0.0070, 0.0150, 0.0393, 0.0150. The eight ROIs plotted correspond to regions for which FDR corrected p-values derived using the group-average (G-1) models were less than 0.05 (see also “Methods” section). A detailed listing of results for all ROIs is provided in Supplementary Table 1. Permutation-based p-values derived from performing RSA at an individual-level revealed significant outcomes ( $p < 0.05$ ) in the following numbers of participants per ROI (indicated in parentheses for personal and group-average models, respectively): L Precun (13,14), R Precun (12,8), L Mid Temp (10,11), L Inf Pariet (7,8), L Post Cing (10,9), L Mid Occ (9,8), L Mid Front (8,5), L Angular (8,7). The cumulative binomial probability of achieving 4 or more outcomes at  $p < 0.05$  in an ROI is 0.04. Source data are provided as a Source Data file.



**Fig. 5 Neuroanatomical distribution of person-specific representational structure (RSA-Searchlight).** Computation of the three RSAs illustrated here, as well as hypothesis testing and FDR correction mirrored the protocol of the above ROI-based analyses translated into a searchlight framework. Differently ROI selection was by passing a 3-voxel radius cube throughout the brain (rather than segmenting anatomical atlas regions). The heat maps illustrate t-statistics corresponding to one sample t-tests of the corresponding RSA coefficients against zero. The t-statistics illustrated correspond to p-values that survived an FDR<sup>64</sup> threshold placed at  $q = 0.1$  ( $q = 0.1$  was used rather than  $q = 0.05$  to enhance the visibility of clusters for display purposes). The anatomical makeup of clusters arising from an FDR threshold of  $q = 0.05$  are listed in detail in Supplementary Tables 3–5. Brain illustrations were made using ref. 103.



**Fig. 6 Individual identity can be decoded from fMRI activity elicited during the imagination of common scenarios.** Each bar illustrates the percentage of times that participant-specific models better predicted the same participant's fMRI representations than another participants' fMRI data (see main text for details). This test was repeated for each pairwise combination of the 26 participants. The eight ROIs illustrated were identified in our initial ROI-based analysis. Complete results for all ROIs are in Supplementary Table 2. To provide context for the neural-decoding accuracy values, we ran a comparative pairwise decoding analysis in absence of fMRI data based on the verbal and attribute models alone (e.g., P1-verbal vs. P1-attribute and so on). This yielded an accuracy of 83% ( $p = 0.0001$ ). Source data are provided as a Source Data file.

We detected decoding accuracies of around 75% across all eight ROIs (all  $p < 0.05$ , see Fig. 6 for precise percentages and permutation test-based  $p$ -values, and Supplementary Table 2 for a complete listing of results for all ROIs). Critically, these results provide evidence that individual identity can be decoded from fMRI activity elicited during the imagination of common scenarios. They also provide an initial estimate of how accurately pairs of individuals can be distinguished based on fMRI data.

**Supplementary analysis overview: results are robust to perturbation of parameters.** In light of the variability in hypothesis testing outcomes that have been observed to arise out of differences in choices of analytic procedure<sup>67</sup>, we conducted a battery of analyses to probe the robustness of the current results to changes in parameterization.

As concerns the anatomical ROI-based RSA we present evidence in Supplementary Fig. 2 that both the verbal and attribute models helped contribute to explaining person-specific fMRI representations (and therefore that multimodal model integration was warranted). In Supplementary Figs. 3 and 4 we demonstrate that the verbal and attribute models both predicted person-specific fMRI representational structure when applied in isolation, and therefore that even though multimodal model combination proved to be beneficial, it was not essential to detect interpersonal differences. In Supplementary Fig. 5 we replicate our main findings when group-average models were differently built by averaging individuals' model representations in model feature space as opposed to similarity space (thus the averaging method was not critical). To demonstrate that results were not tied to the precise number of voxels used in analysis, we replicated similar outcomes when analyzing 50 or 200 informative voxels per ROI (Supplementary Figs. 6 and 7). For completeness, we also include results derived using RSA on generic GloVe representations of the stimulus prompts (GloVe vectors did correlate with the eight ROIs, but cannot account for interpersonal differences, see Supplementary Fig. 8). Lastly, we repeated the RSA when left handers' brain hemispheres were flipped to counteract putative left lateralized language effects, which made little difference (Supplementary Fig. 9).

For the identity decoding analysis (Fig. 6) we generated broadly similar conclusions when analyses were repeated using the verbal and attribute models in isolation (Supplementary Fig. 10) on either 50 or 200 informative voxels per ROI (Supplementary Figs. 11 and 12), and when analyses excluded participants with weak model-to-brain correlations (Supplementary Fig. 13), and when analyses were performed within female/male participant subgroups (Supplementary Figs. 14 and 15).

## Discussion

This study has decoded individual identity from brain activity elicited during the imagination of a set of common situations. Importantly, this has provided evidence that fMRI can measure brain activity with sufficient signal to quantify meaningful interpersonal differences in the neural representation of complex imagined events. In parallel, we have shown that the representational structure of person-specific fMRI signal can be predicted using the current personalized verbal and attribute models. This constitutes a critical advance beyond the one-size-fits-all approach that has typically been applied to predict fMRI data based on group-level models derived from internet data<sup>46–53</sup>. Whilst such group-level models have high signal-to-noise-ratio for capturing population-level commonalities, they cannot distinguish individuals, and they cannot predict person-specific variance across fMRI representations of multiple stimuli as we have done here.

This study revealed that fMRI activation patterns in left/right medial parietal cortex, left temporoparietal junction, left dorsolateral prefrontal cortex, ventral and medial prefrontal cortex, and anterior temporal lobe reflected the representational structure of individual participants' personal descriptions/ratings of their imagined scenarios. This distribution of regions closely resembles the layout of the established core episodic recollection and simulation network<sup>6–8</sup>. Our results are probably best considered to reflect a mixture of recollection and simulation. However, the two are practically difficult to disambiguate here and even some of the scenarios that participants reported as being recollections of real events may have reflected a conflation of information across multiple related episodes<sup>68</sup> (as might be particularly prevalent in our more routine experimental scenarios, such as driving, housework and cooking).

In particular, we found that medial parietal cortex (MPC: precuneus and posterior cingulate) encoded person-specific representations strongly. MPC is routinely activated in episodic recollection/simulation and also in the perception of scenes and space<sup>10–13,69</sup>. MPC is believed to play a role in segmenting events from continuous experience and encoding abstract cross modal event representations during recall<sup>16,17</sup>. Other recent work has revealed that imagined places, contexts and people can all be discriminated from fMRI activation within MPC<sup>19–23</sup> and that MPC — temporal lobe connectivity reflects the content of ongoing thought<sup>70</sup> and memory recall<sup>22</sup>. Our study adds to this by revealing that MPC representations of complex imagined events are person-specific and can be predicted using the current verbal/attribute models.

The left temporoparietal junction (LTPJ: posterior mid temporal cortex, inferior parietal cortex, angular gyrus and mid

occipital gyrus) also strongly encoded person-specific information. Besides sharing a role in event segmentation and simulation<sup>16,17</sup>, LTPJ has been linked to the so-called autonetic<sup>71</sup> conscious subjective experience associated with remembering episodes from a first-person perspective<sup>13,72</sup>, as well as bodily self-consciousness<sup>73</sup>. Indeed, a novelty of the current approach was the initiative to model imagined scenarios from participants' first-person perspectives. However, whilst this approach proved successful, future work will be necessary to evaluate what role perspective played here. For instance, our participants did not additionally describe/rate scenarios from a third-person perspective, so we cannot be sure that such a model would not have worked just as well. We hope to elaborate on this in the future.

Left Dorsolateral prefrontal cortex (DLPFC: mid frontal) also reflected individual differences. DLPFC is typically associated with cognitive control processes that regulate thought and actions according to task goals/personal interests<sup>38,74,75</sup>. Thus, DLPFC is likely to have helped regulate the activation and suppression of scenario memories on cue, possibly by deactivating memory codes in the hippocampus<sup>76,77</sup>. It is interesting (though not unexpected<sup>20</sup>) that DLPFC would encode detailed scenario representations given its putative role in cognitive control. One possibility is that these representations were read from other brain regions to help inform task prioritization. Alternatively, representations could be indirect, reflecting the differential engagement of control processes invoked in the recall/simulation of different modalities of information. Future work exploring the dynamics of information flow between regions may help to clarify this.

The hippocampus, which is a key hub of episodic memory<sup>13</sup> and simulation<sup>2</sup>, was conspicuous by its absence from our results. The hippocampus has been described in terms of providing a spatial scaffolding for episodic memory that indexes the neocortical components that code perceptual, emotional, and conceptual content of experience and provide a sense of autonetic consciousness<sup>13</sup>. In light of this, it is reasonable to suppose that the current attribute/verbal models may better predict the reconstructed neocortical perceptual/emotional/conceptual content than the hippocampal scaffolding. This said, close inspection of our results (Supplementary Tables 1 and 2) reveals that hippocampal (and parahippocampal) ROIs did significantly reflect personal scenario representations prior to correction for multiple comparisons.

In addition, our searchlight analyses identified clusters in: (1) Ventromedial prefrontal cortex, which has recently been found to encode the identity of known people, places and whether they were liked when fictitious meetings are imagined<sup>7</sup>, and more generally is thought to play a role in encoding self-reference<sup>78</sup>, emotional information<sup>79</sup>, and remote (years-old) episodic memories<sup>13,15</sup>. (2) Dorsomedial prefrontal cortex, which has been associated with inferring traits of other people<sup>78</sup>, and has recently been shown to encode collective memories reflecting sociocultural group membership<sup>24</sup>. (3) The anterior temporal lobe, which is thought to be a key semantic memory hub<sup>80</sup>.

It is noteworthy that beyond the anterior temporal lobe, most, if not all of the regions uncovered in our analysis, and more generally most of the core episodic recollection/simulation network<sup>6–8</sup> overlaps with the brain's semantic system<sup>79</sup>. Whilst it has always been acknowledged that episodic and semantic memory share a close relationship, they have classically been distinguished on phenomenological grounds<sup>70,81,82</sup>: Episodic memories are characterized by an awareness of the self, situated somewhere at a unique time and place, whereas semantic memories reflect the sort of general knowledge found in an encyclopedia that lacks a personal space-time context. Since their initial distinction<sup>70</sup>, the

boundaries between the two have become increasingly fuzzy both in terms of theory and also their neural correlates<sup>83,84</sup>. One proposal is that this is because episodic memories bind together familiar semantic concepts (e.g., people, items, places) within an episode-specific sensory and spatiotemporal context<sup>83</sup>. Indeed, this is roughly what our modeling approach has recreated in integrating generic word-level semantic vector representations with scenario-level experiential attributes.

The broader relevance of the success of the current approach is to ground the proposal that it is now practical to study the neural correlates of both episodic and semantic memories within the same unified predictive modeling framework. The verbal and attribute models deployed here were constructed from much the same parts that we<sup>49,52,53</sup> (and others<sup>46–48,50,51</sup>) have used to predict semantic brain activity elicited in language comprehension. In particular, in a separate study, performed on different participants, we have shown how word-level attribute vectors can be integrated with GloVe vectors (both group-level) to predict fMRI activity elicited in (third-party) sentence comprehension. This revealed a “semantic network” of regions that is broadly similar to the episodic simulation network of this study. A difference was that semantic predictions were relatively more accurate in temporal and inferior frontal regions<sup>49,52,53</sup> than in medial parietal cortex (contrary to the case here). However, the key point is that the same modeling basis can be adapted to predict brain activity in the phenomenologically different settings of episodic simulation and language comprehension. A potential advantage of such a unified approach would be to enable features that are believed to differentiate episodic/semantic memory (time/place/autonetic awareness) to be explicitly encoded in a predictive model and evaluated on their ability to predict brain activity elicited in episodic/semantically oriented task contexts.

Indeed, the notion of deploying unified episodic/semantic models could be relevant to other cognitive domains. It has been suggested<sup>84</sup> that episodic and semantic memory systems are anchored on the default network<sup>85</sup>. The default network encompasses multiple interwoven systems<sup>86</sup> that are active in ongoing thoughts<sup>29</sup> such as remembering, envisioning the future and making social inferences. Studies attempting to characterize ongoing thought have typically applied “multidimensional experience sampling”, where participants repeatedly rate the nature of their current thoughts on multiple dimensions<sup>29</sup> (e.g., whether thought is detailed, in images or verbal, related to the past or future...). Not only do within-participant thought samples covary with concurrent default network activity<sup>30,38</sup>, but individual differences in resting-state network activity profiles predict trait-level individual differences in thought profiles sampled outside the scanner at a later occasion<sup>70</sup>, as well as other psychometric traits<sup>37</sup>. Thus, accepting that ongoing thought includes episodic recollection/simulation, we hypothesize that individual differences in the current fMRI measures/models will to an extent reflect individual differences in thought profiles sampled in the wild, and that both measures will combine to predict individual differences in personality and other stable psychometric traits. We also hypothesize that the current neural measures and verbal/attribute models could provide new ways to predict individual differences in ongoing thought at a state-level, tracking how the “meaning” of internal monologues<sup>86,87</sup> and/or the “experience” of daydreaming unfolds over time, and possibly even may forecast how novel future events will be experienced (see also refs. <sup>56,88,89</sup> for image-based computational models of visual imagery).

Beyond basic science, we contend that the current methods could contribute to clinical studies where the ability to detect individual differences is essential at various levels. This could be in helping to characterize and diagnose disorders associated with episodic memory and imagery deficits e.g., Alzheimer's

disease<sup>90,91</sup>, schizophrenia<sup>92</sup>, and depression<sup>93</sup>, and perhaps even personalizing treatments and predicting outcomes (e.g., refs. <sup>94,95</sup>). This study would suggest the current approach is feasible for use in healthy agers at least. How the current results extend to younger adults remains untested. We naturally expect interpersonal differences to also be discernable in younger adults' fMRI data (who are less prone to neural atrophy<sup>96,97</sup>). However, we also anticipate there will be systematic differences associated with socioecological factors, life experience, and neurophysiology<sup>97</sup>. In particular we hypothesize that younger adults' fMRI data will be characterized by the presence of relatively more episodic features, based on behavioral studies showing that older adults' verbal descriptions of past<sup>98</sup> and future-hypothetical<sup>99</sup> scenarios contain more semantic details (not specific in time and place), and fewer episodic details (specific in time and place).

In conclusion, this study has both revealed and predicted meaningful interpersonal differences in fMRI activity elicited as personal experiences were imagined. Firstly, we hope that our approach will help provide a unified modeling foundation for future studies seeking to predict interpersonal differences in the neural correlates of episodic and semantic memory as well as other related cognitive domains. Secondly, we hope that our approach will improve the predictive power of future fMRI studies by demonstrating how person-specific between-stimulus variation can be modeled and explained in terms of model features/scenarios. Finally, we hope that that this work presents a step towards future methods that enable detailed information elicited in mental imagery to be read directly from the brain.

## Methods

The study was approved by the University of Rochester research subject review board (RSRB00067540). All participants were required to understand the experimental procedure and give their consent by signing an informed consent form.

**Participants.** Thirty healthy older adults with normal cognition, indexed by Telephone Interview for Cognitive Status (TICS<sup>TM</sup>)  $\geq 31$ , were recruited for the study. Participants had adequate visual (normal or corrected to normal vision) and auditory acuity for testing, were English-speaking, and community dwelling. Of the 30 participants, 26 yielded viable fMRI data: 2 participants failed to attend fMRI and 2 attended but failed to complete the experiment. Of the 26 participants included in analysis, 5 were left handed and 17 were female. Mean  $\pm$  SD age was  $73 \pm 7$  years old.

**Scenario stimuli.** We identified 20 common scenarios that we anticipated: (1) would have diverse experiential characteristics e.g., whether they are actively or passively experienced, social or asocial, indoors or outdoors etc; and (2) that different participants would have different experiences of e.g., in their degree of activity and social engagement, in venue and so on. This was with the goal of generating sufficient statistical structure in both model and neural data, such that both scenarios and participants would be distinguishable. The number of scenarios was restricted to 20 to meet experimental and scanning time constraints (see also the Experiential attribute model section below). Participants were asked to "vividly imagine themselves in each scenario" prior to providing a brief "sentence-length" description of that scenario. They were later requested to re-imagine the same scenario whilst undergoing fMRI. Scenarios were first read to participants, and later displayed during fMRI in the following form: "A X Scenario", or "An X Scenario" where X is a placeholder for: resting, reading, writing, bathing, cooking, housework, exercising, internet, telephoning, driving, shopping, movie, museum, restaurant, barbecue, party, dancing, wedding, funeral, festival.

**Experiential attribute model.** To characterize participants' imagined experience of the different scenarios we identified 20 experiential attributes. These were selected from a broader set of 65 attributes, which had recently been introduced to model semantic representation using peoples' ratings of their sensory/motor/affective/cognitive experiences with words and their referents<sup>45</sup>. e.g., "Please rate the degree to which your restaurant scenario involves human speech sounds". The 65 attributes were cut down to 20 to meet experimental constraints such that participants could read, imagine, describe and rate the 20 scenarios comfortably within a 2-h time frame. The 20 attributes were selected to broadly span the different domains of experience that were originally identified by Ref. <sup>45</sup>. However, the final choice was ultimately based on the current authors' intuitions of which attributes were most relevant for the current scenarios, as guided by prior experience using the 65

attributes to model fMRI data<sup>49,51,52</sup>. The attributes are listed as follows: bright, color, motion, touch, audition, music, speech, taste, head, upperlimb, lowerlimb, body, path, landmark, time, social, communication, cognition, pleasant, unpleasant. On rating each scenario, the participant was reminded of their personal verbal description of the scenario. A standard description of each attribute was provided to guide ratings, along with examples of words that would receive high and low scores (selected from the original templates introduced by ref. <sup>45</sup>, and listed in detail in Supplementary Table 6). Each attribute rating was normalized within each participant, by subtracting the mean rating for the attribute (across 20 scenarios), and dividing by the standard deviation.

**Verbal model.** To quantitatively model participants' verbal descriptions of their imagined scenarios, we applied a state-of-the-art and freely downloadable distributional semantic model of word-level meaning (GloVe<sup>44</sup>). GloVe represents words as 300 dimensional vectors (see main text in "Results" section for more details). We cut out content words from participants' verbal descriptions of scenarios, and mapped each content word to the corresponding GloVe vector. We then combined word-vectors together to model entire scenario descriptions through pointwise addition.

**fMRI stimulus presentation.** Stimuli were presented on a screen in black Arial font (size 50) on a gray background that participants viewed while in the scanner. fMRI was scanned during a single uninterrupted session in which the 20 scenario stimuli were presented five times over (five runs). Scenario order was randomized within each run. Scenario stimulus prompts (e.g., "A dancing scenario") remained on screen for 7.5 s (3 TRs). The participants had been instructed to re-imagine themselves in the given scenario only when the stimulus prompt was on screen and clear their mind after it disappeared. There was a 7.5 s (3 TR) delay prior to the next scenario presentation, during which time a fixation cross was displayed. Runs were separated by a 15 s interval, in which a second by second countdown was displayed (e.g., "Starting run 2 in 13 s"), which was followed by 7.5 s of fixation cross preceding the first stimulus of the run. The 26 participants who successfully completed scanning all reported that they had been able to imagine the scenarios on prompt.

**MRI data collection parameters.** Imaging data were collected at the Rochester Center for Brain Imaging using a 3 T Siemens Prisma scanner (Erlangen, Germany) equipped with a 32-channel receive-only head coil. The fMRI scan began with a MPRAGE scan (TR/TE = 1400/2344 ms, TI = 702 ms, Flip Angle = 8°, FOV = 256 mm, matrix = 256  $\times$  256 mm, 192 sagittal slices, slice thickness = 1 mm, voxel size 1  $\times$  1  $\times$  1 mm<sup>3</sup>). fMRI data were collected using a gradient echo-planar imaging (EPI) sequence (TR/TE = 2500 ms/30 ms, Flip Angle = 85°, FOV = 256 mm, 90 axial slices, slice thickness = 2 mm, voxel size 2  $\times$  2  $\times$  2 mm<sup>3</sup>, number of volumes = 639).

**MRI preprocessing.** SPM 12 was used to preprocess participants' structural and functional MRI data. Structural scans were co-registered and warped to a common anatomical template in MNI space. The first 10 functional scans during which time steady-state equilibrium was achieved were discarded from the fMRI data. This resulted in the deletion of fMRI volumes associated with the very first (randomly selected) scenario prompt. Thus, a single scenario per participant was represented by 4 rather than 5 fMRI replicates, but otherwise the scenario was treated the same in analyses. The remaining scans were slice-time corrected, motion-corrected, co-registered to their normalized structural images, and then warped to MNI space by applying the same transformation which normalized their structural image. fMRI data was resampled at 3  $\times$  3  $\times$  3 mm<sup>3</sup>.

Six head motion parameters (translation on *x*, *y*, *z* axes, and yaw, pitch and roll, see Supplementary Fig. 18) and linear trend were voxel-wise regressed out from the fMRI data within each of the 5 runs. Specifically, prior to regression, each vector of voxel activation, and each vector of the nuisance parameters (motion/trend) was z-scored within each run by subtracting the vector-wise mean and dividing by the vector-wise standard deviation. A separate multiple regression was computed to predict voxel activation from the nuisance parameters. Finally, the residuals arising from the regression for each voxel were taken forward to use in computing the scenario representations that would form the basis of our subsequent analyses. Individual scenarios were represented by voxel-wise averaging across the four fMRI volumes (of residuals) spanning the period 5 to 15 s after stimulus onset (at 15 s the next stimulus prompt was displayed). The point of onset at 5 s was selected because the hemodynamic response associated with imagination was likely to peak at the very earliest at 5 s post stimulus presentation (the visual response to the written stimulus would be expected to peak at around 5 s, with recall/imagination following later). The interval of 4 volumes were measured to allow for between-participant and between-scenario differences in hemodynamic response profiles. Such hemodynamic differences could arise from a mixture of differences in neural response latencies associated with consciously retrieving scenarios from memory and bringing them to mind, reliving the multisensory experience of the simulated scenario (which could play out for different durations for different scenarios/people), and subsequently suppressing the imagined scenario.

Besides having observed averaging activation to be an effective strategy in other researchers' studies of active thought<sup>35,46</sup> and in our own work on mental simulation/active thought<sup>56,88,100</sup>, we chose the averaging approach in favor of a canonical HRF-based approach because we were not confident that some of the assumptions made by the canonical approach would hold for the current task (which relies on participants to consciously coordinate their recall/imagination, as opposed to say testing for automatic perceptual responses). In particular, the canonical HRF approach would both assume and predict precisely the same neural response profile (including peak response latency and response duration) for each participant and each scenario. This would be modeled by convolving a time series marking the stimulus display periods with a fixed canonical hemodynamic response function. We were initially cautious about this assumption based on our own experience as dedicated participants undertaking different mental simulation studies of a similar ilk, where we have found it difficult to rigorously synchronize our conscious imagination to stimulus onsets and disappearances. To follow up these hunches we performed a preliminary investigation into the degree of interpersonal variation in hemodynamic responses for the current data.

Results of this investigation are included in Supplementary Figs. 19–21. Qualitative inspection suggests that there was indeed substantial interpersonal variability in both the latency and duration of participants' hemodynamic response profiles. For example, some participants appear to have responded rapidly in imagining scenarios and also appear to have cleared their minds relatively sharply. Other participants appear to have responded more slowly and to have kept mental simulations going on for longer. Thus, modeling each participants' activation using the same HRF would have captured some participants' neural response profiles well and others poorly. As the goal of this study was to capture interpersonal variation, we opted for the averaging approach in attempt to avoid biasing our results to reflect only the subset of participants who exhibit canonical responses. We do plan to pursue the nature of these interpersonal (and potentially inter-scenario) differences in neural response dynamics in future work. However, for the time being the current averaging approach provides a practically effective solution to the current problem that lacks the assumption of common peak response latencies and durations across participants. For the future, we note that there may be benefits to including the temporal derivative of the HRF in computation, as was suggested by a reviewer.

**fMRI Voxel selection (within regions of interest).** Because not all fMRI voxels contain informative signal, we estimated which ones were likely to be informative using a commonly used strategy introduced by ref. 46. This identifies stable voxels that had a similar response pattern across each run of 20 scenarios. For each participant, and separately for each voxel, we took each pair of runs, and computed the inter-run Pearson correlation in activation across the 20 scenarios. This resulted in 10 correlation coefficients per voxel (derived through intercorrelating the five runs) that were *r*-to-*z* transformed. A single stability metric for each voxel was derived by taking the mean of the 10 coefficients.

We then parcellated fMRI data into 90 neuroanatomical ROIs using Automated Anatomical Labeling<sup>61,62</sup>. We repeated all of our main analyses on the 100 most stable voxels (marked by high positive coefficients) within each ROI, or all voxels in the ROI if there were fewer than the desired number. Stable voxels were extracted from each ROI, and vectorized. A single representation of each scenario was computed by voxel-wise averaging activation across the 5 scenario replicates. Activation in each voxel was then normalized by subtracting the mean and dividing by the standard deviation of that voxel's activation. To test the robustness of our results to this particular parameterization, we repeated analyses on the 50 and 200 most stable voxels per ROI and obtained similar results, which are reported in Supplementary Information.

**Transformation of personal models and fMRI data to representational similarity space.** To test for a relationship between model and fMRI scenario representations we applied Representational Similarity Analysis (RSA<sup>63</sup>). To set this up, fMRI data and models were re-represented in similarity space. For each of the 26 participants, inter-scenario Pearson correlation matrices were computed separately for the verbal and attribute models (yielding two 20\*20 matrices). All correlation coefficients were *r*-to-*z* transformed ( $\text{arctanh}$ ). The procedure was repeated on each participants' fMRI data, and performed separately on each ROI to generate a total of 90 correlation matrices per participant. The below diagonal matrix triangle (similarity triangle) was segmented from all (model and fMRI) correlation matrices (which are symmetric on the diagonal). Each similarity triangle was then vectorized to form a similarity vector with 190 entries.

**Multimodal similarity model construction.** To construct a multimodal model representation for each participant, their (personal) verbal and attribute similarity vectors were normalized by *z*-scoring, i.e., subtracting the mean value (of the 190 entries) from each entry and dividing each entry by the standard deviation (across the 190 entries). A multimodal representation was computed for each participant by pointwise summation of the verbal and attribute similarity vectors.

**Group-average (G-1) similarity model construction.** To construct group-average model representations that estimate population-level commonalities in scenario

representation, we pointwise averaged model similarity triangles corresponding to the group members. Critically a different group-average model was built for each test participant which excluded that test participant's data. So, in a test of participant 1's fMRI, a group-level model (G-1) was built by averaging model similarity triangles for participants 2 to 26 and so on. Note that group-level representations could equally have been built by averaging individuals in model feature space prior to computing similarity matrices, and this approach yields a similar outcome (Supplementary Fig. 3).

**Representational similarity analyses correlating fMRI data against model data.** To test for a relationship between model and brain data, Spearman's correlation was computed between model and brain similarity vectors. The arising correlation coefficients provided a quantitative measure of the strength of match between fMRI and model representations. The statistical significance of RSA correlations was computed at an individual-level using a standard permutation testing procedure<sup>63</sup>. Scenario order was randomly shuffled, rows and columns of the fMRI correlation matrix (but not model matrix) were rearranged following the shuffled order, the lower matrix triangles of both fMRI and model matrices were extracted, vectorized and then compared to each other using Spearman correlation. This was repeated 1000 times with different shuffles, and the 1000 resulting coefficients were used to build a NULL distribution. Statistical significance was estimated as the fraction of randomly shuffled correlation coefficients that were greater than or equal to the unshuffled correlation coefficient. This produced a *p*-value for each individual participant. For instance, the number of participants with permutation *p*-values < 0.05 is listed in relevant figure captions and indicated by  $i^*$  in Supplementary Information plots.

**ROI selection based on other participants' data.** Because many ROIs were unlikely to represent information associated with the imagined scenarios (i.e., ROIs that are outside the recollection/simulation network), we identified a subset of regions on which to focus our primary individual differences analysis. Our initial interest here was in establishing whether ROIs that represent common group-average scenario structure are further characterized by person-specific idiosyncrasies. This was based on the natural assertion that despite person-specific idiosyncrasies, different people would imagine different scenarios in a broadly similar way (see also ref. 16). We first tested this assertion of cross-participant commonality by performing RSA to compare person-specific multimodal models between each pair of participants. In total this test was performed on 325 pairwise combinations of the 26 individuals. This yielded a mean  $\pm$  SD Spearman's correlation coefficient of  $0.33 \pm 0.11$ , which when *r*-to-*z*-transformed was significantly greater than zero ( $t = 52.8, p < 1e-128$ ). This supported our assertion that there were indeed strong commonalities in scenario representation between individuals.

Then, we isolated ROIs that reflected group-average representations across participants to later test for person-specific information. We ran RSA between each ROI and group-average models (that excluded the test participant) for each test participant. This produced an RSA correlation coefficient for each ROI and each participant. All correlation coefficients were *r*-to-*z* transformed. Then, for each separate ROI, the set of coefficients corresponding to the 26 participants were compared to zero using one sample *t*-tests (1-tailed, anticipating positive correlation). This yielded a vector of 90 *p*-values. The 90 *p*-values were corrected for multiple comparisons according to False Discovery Rate<sup>64</sup> (FDR). FDR corrected *p*-values corresponding to eight ROIs were under the conventional  $p = 0.05$  threshold. These were: Precuneus\_L, Precuneus\_R, Cingulum\_Post\_L, Occipital\_Mid\_L, Parietal\_Inf\_L, Temporal\_Mid\_L, Frontal\_Mid\_L, and Angular\_L. We considered these eight ROIs provisionally as candidates to take forward to the next person-specific test.

However, before continuing forward, we tested whether this selection of ROIs was reasonable from a different perspective. The above ROI selection approach risks overlooking ROIs that might solely correlate with person-specific models (and not the group-average models) and besides, also carries a risk of emphasizing effects associated with the group-average models (if group-average models happen to fit fMRI noise that is not subsequently captured by the personal models). Therefore, we further tested ROI selection based on comparisons between other participants' fMRI data and their personal models. Specifically, to select ROIs for an individual participant (e.g., P1), we computed RSA between every other participant's (e.g., P2 to P26) fMRI data and their corresponding personal models. This was repeated for each ROI. RSA correlation coefficients were *r* to *z* transformed. For each ROI, the set of 25 (*n*-1) coefficients were compared to zero using one sample *t*-tests (one-tailed, anticipating positive correlation). This yielded a vector of 90 *p*-values. The 90 *p*-values were FDR corrected and ROIs with adjusted *p*-values < 0.05 were selected for that participant (e.g., P1). Repeating this procedure across all participants resulted in 7/8 of the previously selected ROIs being identified in every one of the 26 participants. The left-out ROI was Angular\_L which was selected in 20/26 participants. In addition, Frontal\_Med\_Orb\_L, Frontal\_Sup\_L were identified for every single participant, alongside a mixture of other ROIs that differed between participants. To simplify the interpretation and visualization of results we performed our forthcoming analyses on the original fixed set of eight ROIs (rather than a different subset of ROIs for each individual). A comprehensive listing of test statistics for all 90 ROIs

is in Supplementary Table 1. Frontal\_Med\_Orb\_L, Frontal\_Sup\_L are later identified in our searchlight analysis.

**Partial RSA between fMRI and personal models controlling for group-average representations.** To test for idiosyncratic person-specific correlations between personal models and the corresponding participant's fMRI data, we applied Spearman partial correlation. Specifically, partial correlation was computed between each participant's fMRI similarity vector and their personal model similarity vector whilst controlling for the corresponding group-average similarity vector (where the group-average excluded the respective participant). To estimate the statistical significance of partial RSA correlation coefficients at a single individual-level we applied a permutation testing procedure<sup>101,102</sup>. The procedure differs from case of the standard RSA correlation between fMRI and a single model, where it is sufficient to shuffle the fMRI similarity vector as detailed in the previous section. The difference arises because in the case of partial correlation (and multiple regression) shuffling the fMRI vector disrupts its relationship with both the personal similarity vector and the control (group-average) similarity vector, rather than just with the person-specific vector alone (which is desired). To this end, as described in refs. <sup>101,102</sup> the relationship between the control vector and fMRI vector can be regressed out from the fMRI vector to leave a vector of residuals. The residuals can then be shuffled, and the control effect that was just previously regressed out can then be added back onto the shuffled residual vector prior to computing partial correlation.

In more detail, each permutation was evaluated as follows: (A) All values in fMRI, personal and group-average similarity vectors were ranked (as is standard in Spearman correlation). (B) The ranked fMRI similarity vector from A was regressed on the ranked group-average similarity vector. (C) The residuals arising from B were permuted. To accomplish this, the vector of residuals was entered back into the appropriate (triangular) positions of a 20 by 20 (inter-scenario) matrix. The rows and columns of the residual matrix were then shuffled according to a random order (as in RSA). The below diagonal matrix triangle of shuffled residuals was extracted from the matrix and vectorized. (D) The group-average similarity vector was projected back into fMRI similarity space by multiplication with the beta-weights arising from the regression in B. (E) The result from D was added on to the shuffled residual vector from C. (F) The partial correlation was computed between the result from E and the ranked personal model similarity vector, controlling for the ranked group group-average similarity vector.

Steps A to F were repeated 1000 times with different shuffles (in C), and the 1000 resulting coefficients were used to build a NULL distribution. Statistical significance was estimated as the fraction of randomly shuffled correlation coefficients that were greater than or equal to the unshuffled correlation coefficient. The number of participants with permutation  $p$ -values  $< 0.05$  is listed in relevant figure captions and indicated by  $i^*$  in Supplementary Information plots.

**Searchlight RSA/partial RSA analog of anatomical ROI-based analysis.** To build a more precise picture of the neuroanatomical distribution of scenario representations, we replicated the previous ROI analysis under a searchlight framework<sup>65</sup> by passing a cube of voxels (radius 3 voxels, side 7 voxels, mean  $\pm$  SD number of gray matter voxels per cube  $\approx 257 \pm 56$ ) throughout the brain (using the implementation presented in ref. <sup>66</sup>). RSA was computed within each searchlight ROI of each participant's brain, using group-average models and person-specific models. RSA coefficients were  $r$  to  $z$  transformed ( $\arctanh$ ), and assigned to the center of the searchlight ROI. Voxels that were common to all participants were segmented, and one sample  $t$ -tests were applied at each voxel location to test whether RSA coefficients were significantly greater than zero. FDR<sup>64</sup> correction was applied to  $p$ -values corresponding to all voxels (separately for the person-specific test and the group-average test). We then computed partial correlation-based RSA to test fMRI representations against person-specific models whilst controlling for the corresponding group-average models. Partial correlations were computed on the subset of ROIs that were identified as significant ( $p < 0.05$ , post FDR adjustment) in the previous group-average model test and the resulting  $p$ -values were FDR corrected.

**Decoding individual identity from fMRI data.** To further test whether person-specific models could be applied to decode participant identity by identifying the corresponding participants' fMRI data from other peoples', we repurposed an algorithm that was initially introduced to decode word meaning<sup>46</sup>. Two participants were selected at a time, and similarity vectors for the two respective personal models were cross-correlated with fMRI similarity vectors for the two participants. This left four correlation-coefficients associated with both congruent (P1 vs. P1, P2 vs. P2) and incongruent (P2 vs. P1, P1 vs. P2) model vs. fMRI pairings. If fMRI representational structure characterizes individuals, a closer match between an individual's neural data and their own model is expected. To test this, the four correlation coefficients were  $r$ -to- $z$  transformed, then the two congruent model-to-fMRI coefficients were summed and the two incongruent coefficients were summed. If the congruent sum was greater than the incongruent sum, decoding was scored as correct (1), otherwise it was scored as incorrect (0). This pairwise match was repeated for every combination of participant pairs (325), and the overall decoding accuracy was quantified as the mean of the 325 scores. Operating at

random, a 50% success rate would have been expected. The statistical significance of decoding accuracies was estimated using permutation testing: similarity vectors for fMRI data only were randomly shuffled (so that participants became linked to other individual's neural data), and the entire decoding analysis was repeated across all 325 pairs. The random shuffling procedure was repeated 10,000 times to generate a NULL distribution of decoding accuracies, and the statistical significance was estimated as the fraction of shuffled accuracies greater than or equal to the unshuffled accuracy.

**Statistics and reproducibility.** fMRI and behavioral data collection were undertaken a single time. In Supplementary Information we demonstrate the robustness of the current analytic approach to changes in parameterization.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

Source data are provided with this paper. Both preprocessed fMRI and model data to recreate all analyses presented are available at: <https://osf.io/2e7f4>. <https://doi.org/10.17605/OSF.IO/2E7F4>. Please contact the authors for (large) raw fMRI datasets. Source data are provided with this paper.

### Code availability

Matlab v2020a code to recreate all analyses and plot main and supplementary figures is available at: <https://osf.io/2e7f4>.

Received: 26 November 2019; Accepted: 19 October 2020;

Published online: 20 November 2020

### References

- Schacter D. L., Addis D. R. & Buckner R. L. Remembering the past to imagine the future: the prospective brain. *Nat. Rev. Neurosci.* **8**, 657–661 (2007).
- Hassabis, D., Kumaran, D., Vann, S. D. & Maguire, E. A. Patients with hippocampal amnesia cannot imagine new experiences. *Proc. Natl Acad. Sci. USA* **104**, 1726–1731 (2007).
- Hassabis, D., Kumaran, D. & Maguire, E. A. Using imagination to understand the neural basis of episodic memory. *J. Neurosci.* **27**, 14365–14374 (2007).
- Szpunar, K. K., Watson, J. M. & McDermott, K. B. Neural substrates of envisioning the future. *Proc. Natl Acad. Sci. USA* **9**, 104 (2007).
- Schacter, D. L. et al. The future of memory: remembering, imagining, and the brain. *Neuron* **76**, 677–694 (2012). Nov 21.
- Rugg, M. D. & Vilberg, K. L. Brain networks underlying episodic memory retrieval. *Curr. Opin. Neurobiol.* **23**, 255–260 (2013).
- Benoit, R. G. & Schacter, D. L. Specifying the core network supporting episodic simulation and episodic memory by activation likelihood estimation. *Neuropsychologia* **1**, 450–457 (2015).
- Ritchey M., Libby L. A. & Ranganath C. Cortico-hippocampal systems involved in memory and cognition: the PMAT framework. In *Progress in Brain Research*. Vol. 219, 45–64 (Elsevier, 2015).
- Maguire, E. A. Neuroimaging studies of autobiographical event memory. *Philos. Trans. R. Soc. Lond. Ser. B* **356**, 1441–1451 (2001).
- Lundstrom, B. N., Ingvar, M. & Petersson, K. M. The role of precuneus and left inferior frontal cortex during source memory episodic retrieval. *Neuroimage* **27**, 824–834 (2005).
- Fletcher, P. C. et al. The mind's eye—precuneus activation in memory-related imagery. *Neuroimage* **2**, 195–200 (1995).
- Wagner, A. D., Shannon, B. J., Kahn, I. & Buckner, R. L. Parietal lobe contributions to episodic memory retrieval. *Trends Cogn. Sci.* **9**, 445–453 (2005).
- Moscovitch, M., Cabeza, R., Winocur, G. & Nadel, L. Episodic memory and beyond: the hippocampus and neocortex in transformation. *Annu. Rev. Psychol.* **67**, 105–134. (2016).
- Chadwick, M. J., Hassabis, D., Weiskopf, N. & Maguire, E. A. Decoding individual episodic memory traces in the human hippocampus. *Curr. Biol.* **20**, 544–547 (2010).
- Bonnici, H. M. et al. Detecting representations of recent and remote autobiographical memories in vmPFC and hippocampus. *J. Neurosci.* **32**, 16982–16991 (2012).
- Chen, J. et al. Shared memories reveal shared structure in neural activity across individuals. *Nat. Neurosci.* **20**, 115–125. (2017).
- Baldassano, C. et al. Discovering event structure in continuous narrative perception and memory. *Neuron* **95**, 709–721 (2017). Aug 2.

18. Nielson, D. M., Smith, T. A., Sreekumar, V., Dennis, S. & Sederberg, P. B. Human hippocampus represents space and time during retrieval of real-world memories. *Proc. Natl Acad. Sci. USA* **112**, 11078–11083 (2015).
19. Thornton M. A. & Mitchell J. P. Consistent neural activity patterns represent personally familiar people. *J. Cogn. Neurosci.* **29**, 1583–1594 (2017).
20. Robin, J., Buchsbaum, B. R. & Moscovitch, M. The primacy of spatial context in the neural representation of events. *J. Neurosci.* **38**, 2755–2765 (2018).
21. Sreekumar, V., Nielson, D. M., Smith, T. A., Dennis, S. J. & Sederberg, P. B. The experience of vivid autobiographical reminiscence is supported by subjective content representations in the precuneus. *Sci. Rep.* **8**, 1–9 (2018).
22. Silson E. H., Steel A., Kidder A., Gilmore A. W. & Baker C. I. Distinct subdivisions of human medial parietal cortex support recollection of people and places. *Elife*. **8**, e47391 (2019).
23. Benoit, R. G., Paulus, P. C. & Schacter, D. L. Forming attitudes via neural activity supporting affective episodic simulations. *Nat. Commun.* **10**, 1–1 (2019).
24. Gagnepain, P. et al. Collective memory shapes the organization of individual memories in the medial prefrontal cortex. *Nat. Human Behav.* **4**, 189–200 (2020).
25. Dubois, J. & Adolphs, R. Building a science of individual differences from fMRI. *Trends Cogn. Sci.* **20**, 425–443 (2016).
26. Tavor, I. et al. Task-free MRI predicts individual differences in brain activity during task performance. *Science* **352**, 216–220 (2016).
27. Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D. & Kriegeskorte, N. Unique semantic space in the brain of each beholder predicts perceived similarity. *Proc. Natl Acad. Sci. USA* **111**, 14565–14570 (2014).
28. Cetron, J. S. et al. Decoding individual differences in STEM learning from functional MRI data. *Nat. Commun.* **10**, 1–10 (2019).
29. Smallwood, J. & Schooler, J. W. The science of mind wandering: empirically navigating the stream of consciousness. *Annu. Rev. Psychol.* **66**, 3487–3518 (2015).
30. Sormaz, M. et al. Default mode network can support the level of detail in experience during active task states. *Proc. Natl Acad. Sci. USA* **115**, 9318–9323 (2018).
31. Meulenbroek, O., Rijpkema, M., Kessels, R. P., Rikkert, M. G. & Fernández, G. Autobiographical memory retrieval in patients with Alzheimer’s disease. *Neuroimage* **53**, 331–340 (2010).
32. Elman, J. A. et al. Neural compensation in older people with brain amyloid-beta deposition. *Nat. Neurosci.* **17**, 1316–1318 (2014).
33. Maguire, E. A., Kumaran, D., Hassabis, D. & Kopelman, M. D. Autobiographical memory in semantic dementia: a longitudinal fMRI study. *Neuropsychologia* **48**, 123–136 (2010). Jan 1.
34. Addis, D. R., Moscovitch, M. & McAndrews, M. P. Consequences of hippocampal damage across the autobiographical memory network in left temporal lobe epilepsy. *Brain* **130**, 2327–2342 (2007).
35. Just, M. A. et al. Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth. *Nat. Human Behav.* **1**, 911–919 (2017).
36. Mason, R. A. & Just, M. A. Physics instruction induces changes in neural knowledge representation during successive stages of learning. *NeuroImage* **111**, 36–48 (2015).
37. Turnbull, A. et al. The ebb and flow of attention: between-subject variation in intrinsic connectivity and cognition associated with the dynamics of ongoing experience. *NeuroImage* **185**, 286–299 (2019).
38. Turnbull, A. et al. Left dorsolateral prefrontal cortex supports context-dependent prioritisation of off-task thought. *Nat. Commun.* **10**, 1–10 (2019).
39. Landauer, T. & Dumais, S. A solution to Plato’s problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* **104**, 211–240 (1997).
40. Lund, K. & Burgess, C. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Methods Instrum. Comput.* **28**, 203–208 (1996).
41. Cree, G. S. & McRae, K. Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *J. Exp. Psychol. Gen.* **132**, 163–201 (2003).
42. Vinson, D. P., Vigliocco, G., Cappa, S. & Siri, S. The breakdown of semantic knowledge: insights from a statistical model of meaning representation. *Brain Lang.* **86**, 347–365 (2003).
43. Mikolov, T., Chen, K., Corrado, G. & Dean, J. Efficient estimation of word representations in vector space. In *International Conference on Learning Representations (ICLR) Workshop, Scottsdale, AZ. Proc. 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. (New Orleans, Louisiana, 2013).
44. Pennington, J., Socher, R. & Manning, C. D. GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 1532–1543 (Association for Computational Linguistics, Doha, Qatar, 2014).
45. Binder, J. R. et al. Toward a brain-based compositional semantic representation. *Cogn. Neuropsychol.* **33**, 130–174 (2016).
46. Mitchell, T. M. et al. Predicting human brain activity associated with the meaning of nouns. *Science* **320**, 1191–1195 (2008).
47. Wehbe, L. et al. Simultaneously uncovering the patterns of brain regions involved in different story reading subprocesses. *PLoS ONE* **9**, e112575 (2014). Nov 26.
48. Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E. & Gallant, J. L. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532**, 453–458 (2016).
49. Anderson, A. J. et al. Predicting neural activity patterns associated with sentences using a neurobiologically motivated model of semantic representation. *Cereb. Cortex* <https://doi.org/10.1093/cercor/bhw240> (2016).
50. Wang, J., Cherkassky, V. L. & Just, M. A. Predicting the brain activation pattern associated with the propositional content of a sentence: Modeling neural representations of events and states. *Hum. Brain Mapp.* **38**, 4865–4881 (2017).
51. Pereira, F. et al. Toward a universal decoder of linguistic meaning from brain activation. *Nat. Commun.* **9**, 1–13 (2018).
52. Anderson, A. J. et al. Multiple regions of a cortical network commonly encode the meaning of words in multiple grammatical positions of read sentences. *Cereb. Cortex* <https://doi.org/10.1093/bhly110> (2018).
53. Anderson, A. J. et al. An integrated neural decoder of linguistic and experiential meaning. *J. Neurosci.* **39**, 8969–8987 (2019).
54. Paivio, A. *Imagery and Verbal Processes*. (Holt, Rinehart, and Winston, New York, 1971).
55. Andrews, M., Vigliocco, G. & Vinson, D. Integrating experiential and distributional data to learn semantic representations. *Psychol. Rev.* **116**, 463–498 (2009).
56. Anderson, A. J., Bruni, E., Lopopolo, A., Poesio, M. & Baroni, M. Reading visually embodied meaning from the brain: visually grounded computational models decode visual-object mental imagery induced by written text. *Neuroimage* **120**, 309–322 (2015).
57. Herbelot A. & Qasemi Zadeh, B. You and me...in a vector space: modelling individual speakers with distributional semantics. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*. 179–188 (Association for Computational Linguistics, Berlin, Germany, 2016).
58. Mitchell, J. & Lapata, M. Composition in distributional models of semantics. *Cogn. Sci.* **34**, 1388–1429.19 (2010).
59. Kiehl, D. & Clark, S. A systematic study of semantic vector space model parameters. In *Proceedings of the 2nd Workshop on Continuous Vector Space Models and their Compositionality (CVSC) at EAACL*. 21–30 (Association for Computational Linguistics, Gothenburg, Sweden, 2014).
60. Bruni, E., Tran, N. & Baroni, M. Multimodal distributional semantics. *J. Artif. Intell. Res.* **49**, 1–47 (2014).
61. Tzourio-Mazoyer, N. et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**, 273–289 (2002).
62. Maldjian, J. A., Laurienti, P. J., Burdette, J. B. & Kraft, R. A. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage* **19**, 1233–1239 (2003).
63. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis—Connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).
64. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Statistics* **29**, 1165–1188 (2001).
65. Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional brain mapping. *Proc. Natl Acad. Sci. USA* **103**, 3863–3868 (2006).
66. Pereira, F. & Botvinick, M. Simitar: simplified searching of statistically significant similarity structure. In *2013 International Workshop on Pattern Recognition in Neuroimaging*. 1–4 (IEEE, 2013).
67. Botvinik-Nezer, R. et al. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* **20**, 1–7 (2020).
68. Schacter, D. L. & Addis, D. R. The ghosts of past and future. *Nature* **445**, 27–27 (2007).
69. Cavanna, A. E. & Trimble, M. R. The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* **129**, 564–583 (2006).
70. Smallwood, J. et al. Representing representation: Integration between the temporal lobe and the posterior cingulate influences the content and form of spontaneous thought. *PLoS ONE* **11**, e0152272 (2016).
71. Tulving, E. Memory and consciousness. *Can. Psychol.* **26**, 1–12 (1985).
72. Bonnici, H. M., Cheke, L. G., Green, D. A., FitzGerald, T. H. & Simons, J. S. Specifying a causal role for angular gyrus in autobiographical memory. *J. Neurosci.* **38**, 10438–10443 (2018).
73. Blanke, O. Multisensory brain mechanisms of bodily self-consciousness. *Nat. Rev. Neurosci.* **13**, 556–571 (2012).

74. MacDonald, A. W., Cohen, J. D., Stenger, V. A. & Carter, C. S. Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science* **288**, 1835–1838 (2000).
75. Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Ann. Rev. Neurosci.* **24**, 167–202 (2001).
76. Anderson, M. C. & Huddleston, E. Towards a cognitive and neurobiological model of motivated forgetting. In *True and False Recovered Memories*. 53–120 (Springer, New York, 2012).
77. Anderson, M. C. et al. Neural systems underlying the suppression of unwanted memories. *Science* **303**, 232–235 (2004).
78. Van Overwalle, F. Social cognition and the brain: a meta-analysis. *Hum. Brain Mapp.* **30**, 829–858 (2009).
79. Binder, J. R., Desai, R. H., Graves, W. W. & Conant, L. L. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* **19**, 2767–2796 (2009).
80. Ralph, M. A., Jefferies, E., Patterson, K. & Rogers, T. T. The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.* **18**, 42 (2017).
81. Tulving, E. Episodic and semantic memory. In *Organization of Memory* (eds Tulving, E. & Donaldson, W.) 381–403 (Academic Press, 1972).
82. Wheeler, M. A., Stuss, D. T. & Tulving, E. Toward a theory of episodic memory: the frontal lobes and autonoetic consciousness. *Psychol. Bull.* **121**, 331 (1997).
83. Renoult, L., Irish, M., Moscovitch, M. & Rugg, M. D. From knowing to remembering: the semantic–episodic distinction. *Trends Cogn. Sci.* **23**, 1041–1057 (2019).
84. Irish, M. & Vatansever, D. Rethinking the episodic-semantic distinction from a gradient perspective. *Curr. Opin. Behav. Sci.* **32**, 43–49 (2020).
85. Raichle, M. E. et al. A default mode of brain function. *Proc. Natl Acad. Sci. USA* **98**, 676–682 (2001).
86. Buckner, R. L. & DiNicola, L. M. The brain’s default network: updated anatomy, physiology and evolving insights. *Nat. Rev. Neurosci.* **20**, 593–608 (2019).
87. Baird, B., Smallwood, J. & Schooler, J. W. Back to the future: autobiographical planning and the functionality of mind-wandering. *Conscious. cognition* **1**, 20 (2011).
88. Anderson, A. J., Kiela, D., Clark, S. & Poesio, M. Visually grounded and textual semantic models differentially decode brain activity associated with concrete and abstract nouns. *Trans. Assoc. Comput. Linguist.* **5**, 17–30 (2017).
89. Soto, D., Sheikh, U. A., Mei, N. & Santana, R. Decoding and encoding models reveal the role of mental simulation in the brain representation of meaning. *R. Soc. Open Sci.* **7**, 192043 (2020). Mar 18.
90. Greene, J. D., Baddeley, A. D. & Hodges, J. R. Analysis of the episodic memory deficit in early Alzheimer’s disease: evidence from the doors and people test. *Neuropsychologia* **34**, 537–551 (1996).
91. Anderson, A. J. & Lin, F. How pattern information analyses of semantic brain activity elicited in language comprehension could contribute to the early identification of Alzheimer’s disease. *NeuroImage* **22**, 101788 (2019).
92. D’Argembeau, A., Raffard, S. & Van der Linden, M. Remembering the past and imagining the future in schizophrenia. *J. Abnorm. Psychol.* **117**, 247 (2008).
93. Williams, J. M. G. & Scott, J. Autobiographical memory in depression. *Psychological Med.* **18**, 689–695 (1988).
94. Koutsouleris, N. et al. Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. *JAMA Psychiatry* **75**, 1156–1172 (2018).
95. Whitfield-Gabrieli, S. et al. Brain connectomics predict response to treatment in social anxiety disorder. *Mol. Psychiatry* **21**, 680 (2016).
96. Fjell, A. M. et al. One-year brain atrophy evident in healthy aging. *J. Neurosci.* **29**, 15223–15231 (2009).
97. Wyss-Coray, T. Ageing, neurodegeneration and brain rejuvenation. *Nature* **539**, 180 (2016).
98. Levine, B., Svoboda, E., Hay, J. F., Winocur, G. & Moscovitch, M. Aging and autobiographical memory: dissociating episodic from semantic retrieval. *Psychol. Aging* **17**, 677–689 (2002).
99. Addis, D. R., Wong, A. T. & Schacter, D. L. Age-related changes in the episodic simulation of future events. *Psychol. Sci.* **19**, 33–41 (2008).
100. Anderson, A. J., Murphy, B. & Poesio, M. Discriminating taxonomic categories and domains in mental simulations of concepts of varying concreteness. *J. Cogn. Neurosci.* **26**, 658–681 (2014).
101. Freedman, D. & Lane, D. A nonstochastic interpretation of reported significance levels. *J. Bus. Econ. Stat.* **1**, 292–298 (1983).
102. Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M. & Nichols, T. E. Permutation inference for the general linear model. *Neuroimage*. **92**, 15381–15397 (2014).
103. Xia, M., Wang, J. & He Y. BrainNet Viewer: a network visualization tool for human brain connectomics. *PLoS ONE* **8**, e68910 (2013).

## Acknowledgements

This work was partially supported by the University of Rochester CTSA award number UL1 TR002001 from the National Center for Advancing Translational Sciences of the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Author contributions

A.J.A. and F.L. conceived of the experiment. K.M. and A.J.A. collected data. A.J.A. and B. R. analyzed the data. A.J.A. wrote the first draft of the manuscript. D.D.F., K.H., and F.L. contributed to interpreting the results and editing the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-19630-y>.

Correspondence and requests for materials should be addressed to A.J.A.

Peer review information *Nature Communications* thanks Jonathan Smallwood and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020