CHAPTER

27

Representational Similarity Analyses: A Practical Guide for Functional MRI Applications

Halle R. Dimsdale-Zucker[*,§], Charan Ranganath*[*,§]

*Center for Neuroscience, University of California, Davis, CA, United States; §Department of Psychology, University of California, Davis, CA, United States

1. INTRODUCTION

1.1 Research Perspective

Over the past 50 years, we have seen massive progress in our understanding of the neurobiology of memory. Studies in animal models have revealed detailed relationships between neural coding by single cells in the hippocampus that are related to learning and behavior. Neuroimaging—particularly functional magnetic resonance imaging (fMRI)—has enabled researchers to relate global patterns of brain activity to cognitive processes that support memory formation and retrieval. Until relatively recently, it was difficult to bridge the gap between these two approaches because of fundamental differences in the kinds of questions asked in neuroimaging and single-unit neurophysiology.

In a canonical fMRI study of memory, researchers examine whether the magnitude of brain activity during memory encoding or retrieval differs according to whether an item is subsequently remembered or missed (Diana et al., 2007). The relevant question posed by these types of studies is whether "activation" [i.e., blood oxygenation level dependent (BOLD) signal magnitude] is higher when a memory process is successful than when it is unsuccessful. In single-unit recording, researchers also focus on the magnitude of neural activity (i.e., spike rates), but, unlike fMRI, the typical experimental question concerns the selectivity of neurons, rather than the overall amount of activity.

Recent technical advances in neurophysiology have improved the ability of researchers to simultaneously record from large numbers of neurons. With this approach, neurophysiologists have been able to understand to examine what can be decoded from populations of neurons, as opposed to single units (e.g., Leutgeb et al., 2005). In parallel with the rise of population-based analysis approaches in neurophysiology, multivariate analysis approaches have fundamentally changed the kinds of questions posed in fMRI studies of memory. In a typical fMRI data set, activity in any given brain region, such as the hippocampus, will be imaged across a reasonably large number of voxels. Rather than focusing on the mean level of activity across a population of voxels as is done in traditional approaches, multivoxel pattern analysis (MVPA) approaches focus on an examination of patterns of activity across voxels within the population (Haxby et al., 2001; Norman et al., 2006). MVPA has allowed neuroimaging researchers to focus on questions that more strongly parallel the kinds of questions that have been addressed at the neuronal levels in rodents, thereby "connecting the branches of systems neuroscience" (Kriegeskorte et al., 2008a).

In this chapter, we will focus on a specific form of MVPA known as representational similarity analysis (RSA) (Kriegeskorte et al., 2008a). RSA is a technique that is gaining ground as one of the primary data analysis approaches in cognitive neuroscience. Here,

Handbook of In Vivo Neural Plasticity Techniques
https://doi.org/10.1016/B978-0-12-812028-6.00027-6

we will consider how RSA can be used to reveal fundamental insights into how memories are represented in the human brain, go over experimental design for RSA, and cover the pragmatic aspects of how to conduct RSA and how to avoid common analysis and interpretational pitfalls.

1.2 What Is Representational Similarity Analysis, and How Does It Relate to Other Analysis Approaches?

In RSA, the dependent measure is the degree to which voxel patterns are similar across different experimental conditions. Typically, researchers focus on pattern similarity (PS), although some researchers examine the inverse (i.e., dissimilarity) to parallel approaches used in multidimensional scaling metrics (Kriegeskorte et al., 2008a,b). Thus, one would perform RSA and report mean PS along the axis of the resultant summary graph. Although the focus of this chapter is on the application of RSA to human fMRI data, one of the major strengths of RSA is that it is not limited to testing hypotheses from a single data type or species (Kriegeskorte et al., 2008b).

The basics of RSA are rooted in population vector analysis, which has long been a tradition in single-unit recordings (Georgopoulos et al., 1986). The basic idea is that, rather than looking at the mean level of activity as measured with fMRI or with direct neural recordings, one can look at the distributed pattern of activity that is encoded across voxels or neurons. RSA is now beginning to also be applied to electrophysiological (EEG) recordings in humans, (e.g., Kaneshiro et al., 2015), but that is beyond the scope of the present chapter. Later, we will delve into the specifics of how these computations are performed for fMRI data.

When evaluating whether or not to use RSA, it is important to understand what analytic power it provides over analyses of overall signal magnitude or MVPA approaches that relate activity patterns to categorical outcomes (e.g., prediction of whether one is recalling an object or a face). Kriegeskorte et al. (2008a) operationalize these differences as addressing whether the neural response is directly related to a property of the stimulus (first-order isomorphism; more typically the goal of classifier-based multivariate analyses) or whether it is related to understanding the link between the stimulus and its representation (second-order isomorphism; the goal of RSA).

Traditionally, fMRI data have been examined with univariate analyses—so called because the analyses focus on BOLD signal magnitude (i.e., "activation") as the single dependent measure. The relevant question posed by univariate fMRI analysis is: "Does activation magnitude in [Brain area X] differ between [Experimental Condition A] and [Experimental Condition B]." In a mass univariate analysis, the researcher simultaneously runs univariate analyses across every voxel in the brain. For instance, a researcher could test whether most of the voxels in a given brain area show increased activation during successful recall of a studied word as compared with unsuccessful retrieval attempts. This is the expected pattern if this region supports cognitive processes related to episodic memory; in this example, a likely candidate to observe this pattern would be the hippocampus (Diana et al., 2007). A brain area might show changes in overall activity due to processes that are correlated with successful retrieval (e.g., increased engagement of cognitive control or increased efficiency in response selection). Thus, an increase in the amount of activity during recollection of any event does not tell you anything about whether a brain area represents anything about specific events that are recollected. The latter question is what memory researchers are typically most concerned with, and where MVPA approaches such as RSA can be helpful.

Pattern classification approaches begin to solve this problem of utilizing voxel-by-voxel variability to understand cognition. Pattern classification relies on the idea that there are systematic differences in the variability of voxels that differ between conditions of interest. For example, if individuals had studied either objects or faces, the idea is that some proportion of voxels are preferentially engaged when viewing a house, whereas others are engaged when viewing a face. Intriguingly, these voxels may even be in the same region (Rhodes et al., 2004). This systematic difference in which voxels are engaged, or the magnitude of their engagement, can be used by a classifier to predict whether on any given trial an individual was viewing a house or an object (Haxby et al., 2001).

MVPA, which is a type of pattern classification, has been extended beyond simple classification of object categories to answer questions about topics as diverse as binocular rivalry (Haynes and Rees, 2005) to lie detection (Davatzikos et al., 2005). Different classification schemes can be used, including linear classifiers, neural networks, linear support vector machines, and Gaussian Naïve Bayes classifiers (for a review and suggestions of which method to use, see Norman et al., 2006). Regardless of these implementation details, a huge advancement of MVPA over univariate analyses is the ability to leverage the variability within a region to understand the coding scheme of the brain. MVPA is most powerful when looking for the presence or absence of a given category [e.g., object/face (Haxby et al., 2001)] or cognitive state [e.g., the category of information that is about to be remembered (Polyn et al., 2005)]. Where it falls short is in flexibly thinking about shared and dissimilar features between items across a high-dimensional space (e.g., classification is not optimal for addressing whether

a brain region maps a continuous dimension of experience, such as relationships between events in space and time).

RSA is optimal for looking at a high-dimensional representational space where items can be related to one another in various ways, and it is particularly well suited for continuous, rather than discrete, relationships between items (Kriegeskorte et al., 2008a). For instance, in one of the first applications of RSA, Kriegeskorte showed that it could be used to help understand and test different models of visual representations across species (Kriegeskorte et al., 2008b). Because the goal of RSA is to test whether or not the observed (neural) representation matches an idealized or theoretical representation, one can essentially test any reasonable hypothesis with the data and can even be used in an exploratory manner to help understand how the data are structured. In the following sections, we will discuss practical issues for designing and analyzing with RSA and will talk about these different model comparison approaches in depth.

2. HOW TO DO REPRESENTATIONAL SIMILARITY ANALYSIS

2.1 Overview

RSA is not a complicated method, but, as with any method, one can optimize experimental design and data analysis to make best use of the method. In this section, we will describe the issues to consider, starting from experimental design to interpretation of results. In brief, a task will need to be designed such that trials can be isolated from one another, pairs of trials from different conditions can be related to one another with a similarity metric (e.g., correlation), and, finally, comparisons between the summary values of these condition-wise similarity metrics are made and conclusions drawn. The details of these steps follow.

2.2 Moving From Univariate to Multivariate

One of the things that sets RSA apart from more traditional univariate analyses is that all trials are recombined based on their relationships to one another. Thus, one might compare all trials in one condition against all trials in another condition in a univariate analysis, whereas one might correlate trials within the same experimental condition and compare correlations between trials that share a particular feature and trials that do not share this feature in a multivariate analysis. To illustrate this further, let us continue with our example of a memory paradigm where participants are scanned while performing a recognition memory test using a real example (Dimsdale-Zucker et al., 2018). A typical memory-related univariate contrast might compare activity between studied items that were correctly remembered during the scan from studied items that were forgotten. This analysis can be of great value in identifying regions [e.g., the hippocampus (Diana et al., 2007)] or processes [e.g., presence/absence of recollection (Yonelinas, 2002)] that track memory for the items. However, one could ask different questions—for instance, when recollection occurs, does a brain region of interest (ROI) represent information from the original episode?

To answer this question, we designed a study where participants studied objects in two different virtual reality homes. Participants studied sets of 10 objects within one of these two homes across a series of 20 videos. Thus, each object had a unique spatial (house) and episodic (video) context (Fig. 27.1). At retrieval, participants were simply asked to make a memory judgment about studied and unstudied objects while in the MRI scanner. To understand how spatial and episodic contextual information from encoding was represented at retrieval, we used RSA.

To address this question with RSA, we can test whether items that were associated with the same encoding context share neural similarity when they are retrieved. In theory, this question could be addressed by RSA if one is willing to assume that voxel patterns can reflect distributed neural representations in the hippocampus (or in any other brain region, for that matter). However, even if this assumption holds, RSA does not necessarily reflect representations that reflect the goals of the experimenter (Stark et al., 2017). It is here that sound experimental design is especially important.

Here, because, we were interested in understanding whether having the same or different encoding context influences the neural pattern observed at retrieval, we compared the activity pattern elicited during retrieval of items that were each associated with the same study context (e.g., the umbrella and tent occurred in BrownHouse/Video1 (Dimsdale-Zucker et al., 2018)). Likewise, we compared the patterns elicited during retrieval of the other "same context" trial pairs (e.g., umbrella/dumbbells, umbrella/bean bag, umbrella/chandelier, tent/dumbbells, etc. in BrownHouse/Video1 and air plane/football helmet, airplane/telephone booth, etc. in GrayHouse/Video1). As we will describe in the following section, these correlations, in isolation, are difficult to interpret without a baseline for comparison. The most obvious comparison would be to use the same trials but this time examine correlations between all possible "different context," trial pairs—that is, trials were not associated with the same encoding context (e.g., umbrella/football helmet, umbrella/telephone booth,

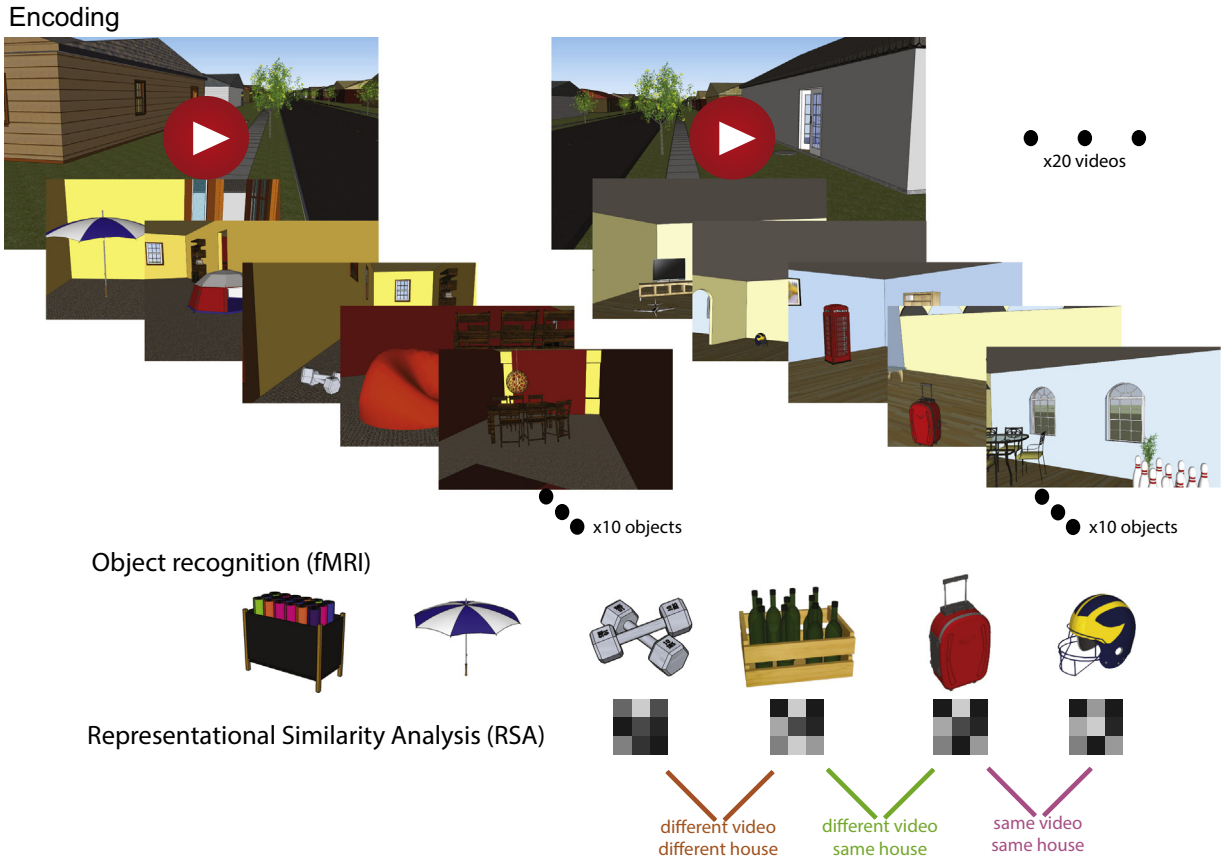

FIGURE 27.1 Experimental design from Dimsdale-Zucker et al. (2018). Participants encoded objects uniquely located within one of two virtual reality homes (spatial contexts) across a series of 20 videos (episodic contexts). Next, they were scanned while performing an object recognition test, which required differentiating old and new objects presented without any contextual information. We used representational similarity analyses to examine the similarity of voxel patterns elicited by each recollected object relative to other recollected objects that were studied in the same (or different) spatial and episodic contexts. *fMRI*, functional magnetic resonance imaging. *Figure adapted with permission of author.*

tent/football helmet, etc.) To reiterate, the same trials are included in each bin of the analysis, which is a fundamental difference between RSA and univariate analyses.

2.3 How Many Trials Do I Need?

As a general rule, an experiment should be designed such that one can get as many trial pairs as possible for the relevant experimental question. What is the optimal number of trials? More is almost always the answer, and the optimal number will depend of course on the underlying effect size. As a general rule of thumb, one might aim for at least 30 trial pairs in the smallest bin.

In the abovementioned example, it should be clear that there will be more trial pairs in the different context bins than in the same context bin. Why could this present a problem? We know that having more trial pairs should yield a better estimate of the true correlation value (see Section 2.8 for how these correlations are computed), especially because having noisy trials can have a more significant influence on the correlations with low trial numbers. Thus, one might be concerned that the observed patterns between the conditions may differ simply because of the differences in trial pairs between them if there is more instability in estimating the patterns from the bin with fewer trials. In practice, we have not seen this to be the case when one is averaging across a reasonably large number of trial pairs.

If unequal bin sizes between conditions is a concern, one option is to randomly resample the condition with the smaller number of trial pairs to match the number of trial pairs between conditions. You can see an example of this in Fig. 27.2 using data from a real study conducted in our laboratory (Dimsdale-Zucker et al., 2018).

Another complication, in answering how many trials are needed, is how one plans to analyze the data (i.e., within a subject and within run, within a subject and

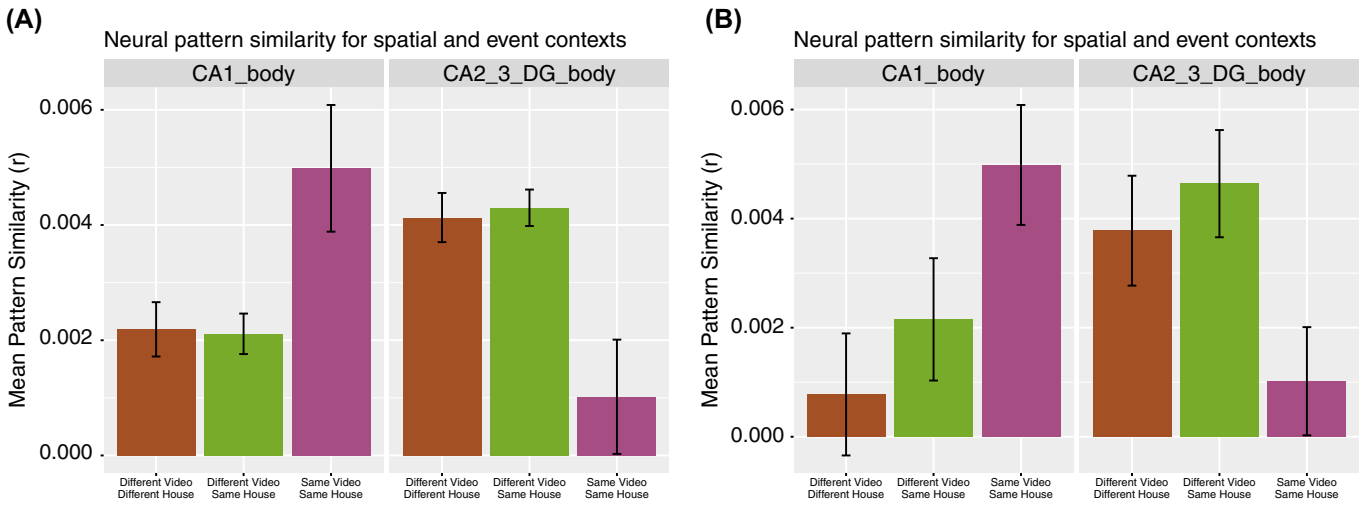

FIGURE 27.2 (A) Mean pattern similarity (PS) including all possible trial pairs in each condition. Error bars represent standard error of the mean. (B) Mean PS where each subject's data have been resampled such that each condition has the same number of trial pairs. Although the data in panel B are noisier (e.g., larger error bars), the overall general pattern of interest (greater PS for items that occurred in both the same video and house in CA1 compared with reduced PS for these trials in CA2/3/DG) remains the same. *Figure adapted from Dimsdale-Zucker et al. (2018) with permission of author.*

between runs, across subjects). We will return to these issues in Section 2.5.

2.4 How Should Trials Be Spaced?

We will cover issues related to modeling of both RSA and other fMRI designs in Section 2.5, but let us first address the issue of trial spacing and randomization because these both impact how well we are able to uniquely account for the variance associated with each trial.

A large problem for RSA, and fMRI in general, is the inherently correlated nature of the data. This is because, even in the absence of a cognitive process, there are ongoing biological and mechanical fluctuations in the signal that can be measured with MRI. To reduce the overlap in the signal that is measured at two time points, the two time points need to be reasonably far apart such that the variance accounted for by each time point can be uniquely estimated by the regression model. Classically, this meant long spacing—on the order of tens of seconds—between trials [slow event-related designs (Blamire et al., 1992)]. Later work revealed that by randomizing spacing between trials within a given range of both long and short values (i.e., "jittering" the inter-trial interval (ITI); Burock et al., 1998; Clark et al., 1998; Dale and Buckner, 1997) one can more efficiently estimate activation across trials within a particular experimental condition. Such fast event-related designs are now standard in most modern fMRI studies. However, for RSA, because we are modeling each trial in isolation, we want each trial to be maximally isolated from all other trials.

In a jittered design, some trials would occur closer in time and others further apart, which is ideal when your model can capitalize on the variance in time across trials within a condition, but it is not ideal for isolating trial-unique activity. Zeithamova et al. (2017) systematically compared the timing of stimuli on how well they were able to use RSA to understand category-level differences, item-level differences, and memory-related differences across a set of canonical regions of interest known to be sensitive to these features. Across the durations they tested, both item-level and memory-related differences were more robust with longer intervals between stimuli (the longest stimulus onset asynchrony they used was 12 s) consistent with extant evidence that longer trial spacing may be better for estimation of single-trial activity patterns (Visser et al., 2016); the ability to detect shared category-level information did not vary by timing condition. Jittering the onset timing of trials made a minimal difference in their ability to detect different representations. Thus, it seems that for RSA designs with many conditions, it is optimal to maximally separate trials from one another rather than trying to orthogonalize at the condition level as would be carried out for univariate designs, although this issue is perhaps more critical for within-run than between-run comparisons. It is critical to strike a balance between what is best for the purposes of modeling and what makes the most sense psychologically to participants. As Mumford et al. (2014) has already written extensively about, the spacing and order of trials should be run- and subject unique to minimize the possibility of false positives.

2.5 How Should I Model my Data?

2.5.1 Within-Run Versus Between-Run Modeling

Once the task has been optimized and the data collected, there are a series of choices that are made at analysis that can impact the interpretation of the results. One that is often overlooked (see Mumford et al., 2014) is whether you plan to compare voxel patterns across trials that occurred within the same run or trials that occurred in different runs. Either approach is reasonable, but correlating voxel patterns across trials within the same run and across trials in different runs is not appropriate.

Although we would like to assume the BOLD response is only driven by neural influences, we know that there are ongoing biological and mechanical fluctuations that have a relatively slow drift over time when compared with the time course of cognitive processes (Heeger and Ress, 2002). This necessarily means that neighboring trials share more variance, even if it is "uninteresting" variance that is unrelated to the task. Not only do neighboring trials share more variance, but trials within the same run are more similar to one another than trials in other runs again due to task-invariant reasons such as scanner drift. This means that if both within- and between-run trials were included when comparing PS, correlation values are being artificially inflated. PS should be higher for conditions that include more correlations between trials in the same run than for conditions that include more correlations between trials in different runs[1].

The issue of whether to look at within-run PS or between-run PS is often determined by the task. In a memory design, or other paradigms where you may have little control over how many trials are in each bin, it can often be preferable to do only between-run RSA because this should increase the total number of trial pairs in each condition. However, for some designs, this does not make sense. For example, Jonker et al. (2018) had a design where they were interested in comparing the overlap in RSA between items at encoding and items during practice (either a restudy or a retrieval condition). In this design, to control for the time between encoding and practice, all items were encoded and practiced in the same run. Thus, RSA necessarily had to be performed within the same run.

A related issue is whether it is better to have fewer, longer runs or more, shorter runs. For other classification techniques such as MVPA where one run is often left out to verify the classification, it is advantageous to have more runs (Coutanche and Thompson-Schill, 2012). If you plan to do within-run RSA, longer runs are better because this means that you can have more observations in each run. For between-run RSA, shorter runs can also be preferable. Our standard practice is to do what makes the most sense for the cognitive processes in the task and to achieve the cleanest data possible (e.g., runs that are too long often have more motion artifact).

2.5.2 Preprocessing

Whether the RSA will be computed between trial pairs in the same run or across runs, the data cleaning and preprocessing procedures are similar. Indeed, the initial preprocessing stages for RSA are nearly identical to those for standard univariate analyses (for a set of standardized recommendations for general preprocessing, see https://fmriprep.readthedocs.io/en/stable/). In our laboratory, we first start by running some type of data quality check (e.g., http://cibsr.stanford.edu/tools/human-brain-project/artrepair-software.html) which generates information about general noisy time points that can be included as spike regressors in our fMRI model and motion parameters to again use in our fMRI model. We then typically coregister the data within each functional run, realign the functional data into structural space, and minimally smooth the data (e.g., with standard 3 mm isotropic voxels, we might use a 2 mm full-width at half maximum (FWHM) smoothing kernel). Although it may seem odd to smooth across voxels if we plan to look at a voxel-wise pattern, a moderate degree of spatial smoothing can enhance the signal-to-noise ratio and enhance classification accuracy (Op de Beeck, 2010) while preserving distributed pattern information (Kriegeskorte et al., 2010). If you are looking at voxel patterns in regions where anatomical precision is integral to the hypothesis (e.g., differences in voxel patterns between subfields of the hippocampus), no smoothing should be performed.

2.5.3 Single-Trial Modeling

One important consideration for RSA is how the data are to be modeled. This is a potential concern because, due to the sluggishness of the BOLD signal, one must deconvolve overlapping responses to adjacent, closely spaced trials. This is usually not a significant concern in standard univariate fMRI analyses, in which one usually wants to estimate the average response across all of

[1] It is conceivable that this confound could be avoided if you designed your task such that there are equal numbers of all conditions in all runs, but, in practice, this is usually impossible with a memory experiment (and other designs that sort trials based on participants' responses) in which trials are sorted based on subject performance. Moreover, it is still not advisable to incorporate a systematic source of unmodeled variance into any statistical analysis.

the trials from a given condition in the run (Burock et al., 1998; Clark et al., 1998; Dale and Buckner, 1997).

In some experiments, one can model the data using essentially the same approach that is used in a standard univariate analysis. For instance, if there are multiple repetitions of the condition of interest, and if you are computing pattern correlations across runs or across subjects, it can be advantageous to model the data with a regressor that models all of the trials within the condition of interest. This approach is generally optimal because it is straightforward to obtain stable estimates of activity for each condition because the model is fit to multiple observations. However, this approach might not be optimal for many memory paradigms. For instance, if you are presenting studied and novel images at retrieval, you do not want to repeatedly probe individuals' memory for each item. Even if there are repetitions of items within a condition (e.g., you could model all remembered trials together even if these are all different items), if you want to combine trial pairs across different conditions (e.g., remembered trials may have been associated with different encoding contexts), in which case, it may not be appropriate to model these trials together.

For cases in which it is not feasible to use a traditional design and analysis approach (i.e., jittering ITI and randomizing order of conditions), Mumford and colleagues have written about several ways to estimate responses to each trial in isolation (Abdulrahman and Henson, 2016; Mumford et al., 2012, 2014, 2015). These approaches essentially involve running a different generalized linear model (GLM) analysis for each trial that is to be estimated. In each model, one trial of interest is modeled separately from other trials. There are different variants that differ in how the trials of noninterest are modeled to make them maximally orthogonal to the trial of current interest. In practice, there should not be drastic differences between the different modeling approaches (Zeithamova et al., 2017), but it is worth choosing the specific modeling approach that is most appropriate given your experimental question.

Even with the single-trial modeling approaches recommended by Mumford, there are relatively strong correlations in estimated responses across trials that are trials closely spaced in time (see Fig. 27.3 for an example). Plotting a correlation matrix and looking to see where the band of autocorrelation drops off can be invaluable in determining how many steps between trial pairs to remove.

Some have also suggested including the first-order temporal derivative as a way of soaking up the additional hemodynamic differences with varying lags between trials (Calhoun et al., 2004; Friston et al., 1998; Worsley and Taylor, 2006). However, when trials are relatively close in time, these temporal derivatives can actually contribute to colinearity between regressors in the model. Therefore, for single-trial modeling approaches, it is generally advised to not include the temporal derivative (Abdulrahman and Henson, 2016; Mumford et al., 2012).

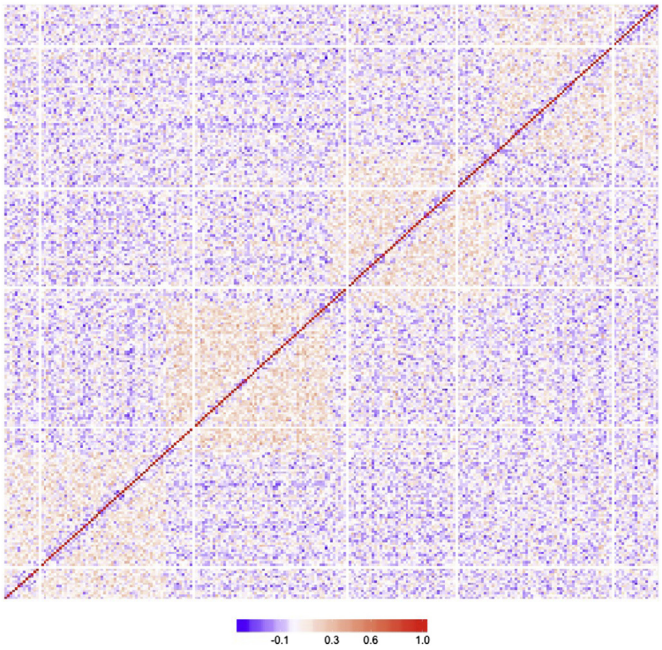

FIGURE 27.3 Graphical representation of all trial pairs for a single subject across four runs [for full data set, see Dimsdale-Zucker et al. (2018)]. Warmer colors (red) indicate higher pattern similarity (PS) (Pearson's r); white lines indicate excluded trials (e.g., poor behavioral performance, outlier beta value). Several patterns in the data are quite striking: First, we observe the bright red diagonal; this is each trial's correlation with itself, which therefore has a value of 1. Moving just off of this diagonal, we can observe a band of approximately 4–5 trials that have lower correlations than all other trial pairs within each run. These trials should be removed from analysis if computing within-run PS because they are contaminated with autocorrelation. Finally, we can also see an overall trend for higher PS within the same run than between runs (notice that the rectangles for each of the four runs are warmer in color than any of the between-run PS values).

Regardless of how you choose to model your data, by the end of this stage, the goal is to have a single beta image (i.e., a matrix of parameter estimates for each voxel in the imaged region) for each trial, or trial type, of interest. These beta images will be used to compute PS values (see trial-level cleaning for an argument for using t-maps instead of betas for RSA).

2.6 How Do I Use Representational Similarity Analysis to Test a Hypothesis About Specific Regions?

The goal of RSA and other multivariate techniques is to understand how information is represented. Often the goal is to test a principled hypothesis about how a ROI

in particular represents information. To continue with the example from our laboratory introduced earlier, we used an ROI-based RSA approach to ask how different subfields of the hippocampus represent contextual information when items are retrieved (Dimsdale-Zucker et al., 2018). To do this, we measured activity patterns in subfields CA1 and CA2/3/dentate gyrus (DG) when participants remembered items that had earlier been studied relative to different spatial (virtual homes) and episodic (videos that occurred in these homes) contexts. We found that an item's spatial context did not distinguish the pattern of representation in the subfields. However, when we looked at items that either had occurred in the same or across different episodic contexts, we found different patterns in CA1 and CA2/3/DG; in CA1, items that shared an episodic context were represented as more similar than those from different episodic contexts, whereas CA2/3/DG showed the opposite pattern and differentiated between objects that had occurred in the same episodic context. Taking an ROI-based RSA approach thus allowed us to answer a question about how different subfields represent contextual information and adjudicate between competing hypotheses in the literature.

ROIs are typically delineated on the basis of anatomical features (e.g., cortical delineations identified with FreeSurfer), task-responsiveness (e.g., regions affiliated with a keyword on neurosynth.org), or connectivity profiles (e.g., regions that are preferentially connected to a node in a network). The only major issue to avoid is that an ROI should not be defined in a manner that is redundant with the PS analysis (Kriegeskorte et al., 2009). For instance, if you wanted to test whether items that are remembered are more similar to one another than items that are forgotten, you should not define ROIs on the basis of a univariate contrast between remembered and forgotten items.

Once the ROIs have been selected, it is necessary to extract activation estimates for each voxel in the ROI separately for each trial or condition of interest. For any given trial or condition, the corresponding voxel pattern vector is comprised of the collective set of activation estimates across the voxels. The voxel pattern vectors across all the trials can be concatenated in a voxels-by-trials matrix. The dimensions of this matrix will vary from ROI to ROI because ROIs are defined by anatomy and not absolute number of voxels. We can next compute a similarity metric (e.g., Pearson's r) across each pair of voxel pattern vectors which will yield a square, trial-by-trial similarity matrix (see Section 2.8 for other approaches to computing trial-wise similarity). If you are doing between-run correlations, this would be a trial-by-trial matrix for all trials across all runs.

On a procedural note, it is critically important to understand which trials fall into which conditions. In our laboratory, we have found it handy to have a separate matrix that shares the trial-wise dimension of the trial-by-trial correlation matrix across either rows or columns (typically, trials as rows has been our laboratory's convention) and where all of the other potentially relevant trial information (e.g., the trial's position within encoding and within a run at retrieval, run membership, the subject's response both at encoding and at retrieval, any encoding manipulation of interest, and any other task-relevant trial features) is contained in the other dimension. This trial information matrix can then be used to index the trial pairs that should be extracted for various conditions of interest.

2.7 How Do I Use Representational Similarity Analysis to Test a Hypothesis That Is Not Regionally Specific?

If you wanted to call Batman for help, you could illuminate the entire sky over the affected city—this is akin to the ROI approach. Or, one could systematically move a spotlight across all points in that given area—this is the general idea behind a searchlight. Either should result in Batman coming to the scene of the crime, or, in our case, understanding the pattern within a ROI, but the size of the search space differs. Searchlight MVPA analyses (Kriegeskorte et al., 2006) are typically performed in cases where one does not have a strong hypothesis about the region that will show the expected effect. With a searchlight, instead of looking across an entire anatomical region, you are instead creating a small volume that is moved throughout either the whole brain or throughout a particular ROI. For this reason, the size and shape of the searchlight can have large consequences on the observed effects. Before we tackle the issue of selecting a shape and size, it is important to understand that the dimensions of the searchlight determine the smoothness of the pattern you are trying to find—a small searchlight is more likely to identify small, blocky patterns, whereas a larger searchlight is smoothing out the pattern that you are looking for (Etzel et al., 2013; Viswanathan et al., 2012).

In our laboratory, we are typically interested in looking at memory-related activity either in the hippocampus itself or in the surrounding medial temporal cortical regions. For this reason, our laboratory has favored relatively small searchlights with a diameter of approximately five voxels and that have a central node with arms of equal length (two voxels out from this central node) extending in three dimensions [see Fig. 27.4; for a published result with a searchlight like this, see Libby et al. (2014)]. However, others have used spherical, diamond, or other shaped searchlights (for implementations of some of these shapes, see http://cosmomvpa.org/). We then treat this searchlight volume

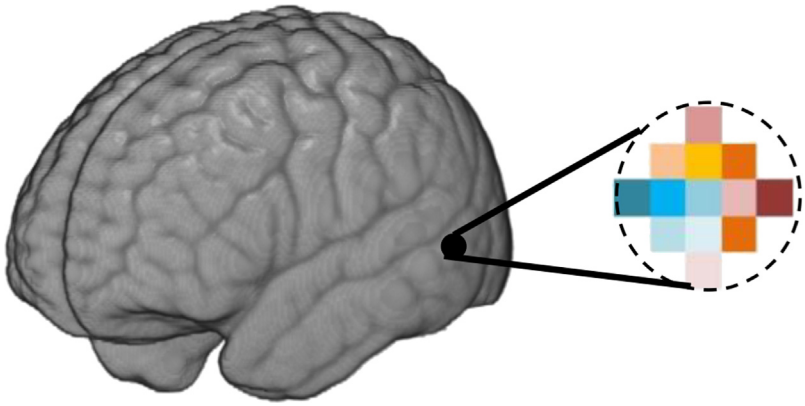

FIGURE 27.4 Example searchlight with a diameter of five voxels. *Figure courtesy of Alex Clarke. For data published with this searchlight, see Clarke et al. (2016).*

as if it were a tiny ROI and extract the values from all voxels within that searchlight volume, repeat this procedure across all trials for this searchlight volume, and eventually compute a pattern similarity value for that searchlight that is assigned to the central voxel node. Next, one can systematically move the searchlight mask across the brain, computing a voxel pattern similarity analysis at each position. This procedure is repeated until all locations within the search space have been sampled. We exclude any section of the searchlight that extends beyond the volume of interest, and, if the total number of voxels in the searchlight space is ever less than nine voxels, we do not compute a pattern correlation for this central node location. This cut-off ensures that estimated correlations will be relatively stable.

Unlike with ROI-based RSA where activity for pairs of trials are extracted, with a searchlight it is more common to subtract the searchlight maps for two conditions of interest. This yields maps for the entire searchlight space where there is a difference in the voxel-wise pattern in the searchlights. Although these maps can look like univariate contrast maps, their interpretation is different (Etzel et al., 2013). It is more appropriate to talk about a searchlight that is centered in a region rather than to talk about the region itself as being implicated in the representation (Etzel et al., 2013).

2.8 How Do I Quantify Similarity?

As suggested by the name, RSA is about analyzing similarity between patterns of activity evoked during different conditions. This can be a tricky issue, however, because similarity can be defined and measured in different ways. If we think about representing trial similarity values in the simplest way possible, we can think about noting each trial's value as a point in x/y coordinate space and extending lines (pattern vectors) joining the origin to these points. We can then think about summarizing the distance between these two points either relative to a line connecting them (think about drawing the third arm of a triangle) or as the angle between these pattern vectors as measured at the origin (for an example, see Bobadilla Suarez and Love, 2017; Walther et al., 2015). Metrics that are similar to thinking about the line that would complete a triangle are distance metrics, such as Euclidean distance and Mahalanobis distance, and metrics that use the angle are Pearson's correlation and cosine distances. Anything that shifts the distance to the origin (e.g., shifts in the MRI baseline) impact angle-related but not line-related metrics, and changes in the length of the lines effect line-related but not angle-related metrics. We will start with a description of Pearson's correlation distance because it is probably the most commonly used distance metric in RSA.

Pearson's correlation was used in one of the first applications of RSA, and it is the most commonly used similarity metric in current neuroimaging studies (Aguirre, 2007; Haxby et al., 2001; Kiani et al., 2007; Kriegeskorte et al., 2008b). Pearson's correlation has the advantage of being straightforward to compute and it is, mathematically speaking, invariant to differences in the mean and variability of the data (Kriegeskorte et al., 2008a). Cosine distance is closely related to Pearson's correlation in that it is concerned with the angle of the pattern vectors. In fact, it is identical to Pearson's correlation in the case where the mean pattern has been removed from each voxel (also called cocktail-blank removal; Misaki et al., 2010; Walther et al., 2015). In practice, many readers are unfamiliar with cosine distance, whereas Pearson's correlation is well known. However, Pearson's correlation is more often used as a measure of association between two variables, rather than a similarity metric, which can lead to some intuitive misinterpretations. For instance, PS between two trials could, in theory, be associated with an r value of −0.8. Intuitively, this would imply a strong association

between two variables, but, in the context of RSA, the negative correlation means that there is very low similarity between the voxel patterns. This interpretive trap can be avoided by plotting the correlation distance metric, $1 - r$, which rescales the values from between 0 and +2.

An alternative to angle-related measures are Euclidian distance metrics (Edelman et al., 1998). Euclidian distance is a "true" distance metric—that is, it is based on the Euclidian geometric system and is interpreted as the distance between two points in a multidimensional space (Walther et al., 2015). Mahalanobis distance is closely related but quantifies the distance between points as a function of standard deviations from the mean (Kriegeskorte et al., 2006). In essence, Mahalanobis distance takes the covariance structure of the entire data set into account, whereas Euclidean distance can be computed without estimating covariance.

Tests of the reliability of these various distance metrics have not revealed systematic differences (although for a more thorough discussion of when there are differences, see Walther et al., 2015) and RSA toolboxes (e.g., CoSMoMVPA, http://cosmomvpa.org/) allow the use of different distance metrics usually by changing one line of code. We therefore encourage researchers to pick the metric that is most appropriate for the question(s) of interest and to consider what has been used in similar studies.

2.9 How Can I Test Hypotheses Using Voxel Pattern Information?

Having computed PS values for each condition or trial bin of interest, these data can be used to test hypotheses. There are at least two major ways to use PS values to summarize the trial-wise patterns for RSA. One approach is to compose a theoretical correlation matrix, or expectancy matrix, that one thinks summarizes the relationships between conditions (see Fig. 27.5C–E for an example Clarke et al., 2016; Kriegeskorte et al., 2008a). This approach is often useful when the relationships between conditions are complex or to compare the observed data against the theoretical predictions of different computational models. The observed trial-wise RSA matrix is then correlated with the expectancy matrix to obtain what is called a second-order similarity matrix. Currently, best practice is to use Kendall's Tau to evaluate second-order similarity (Nili et al., 2014).

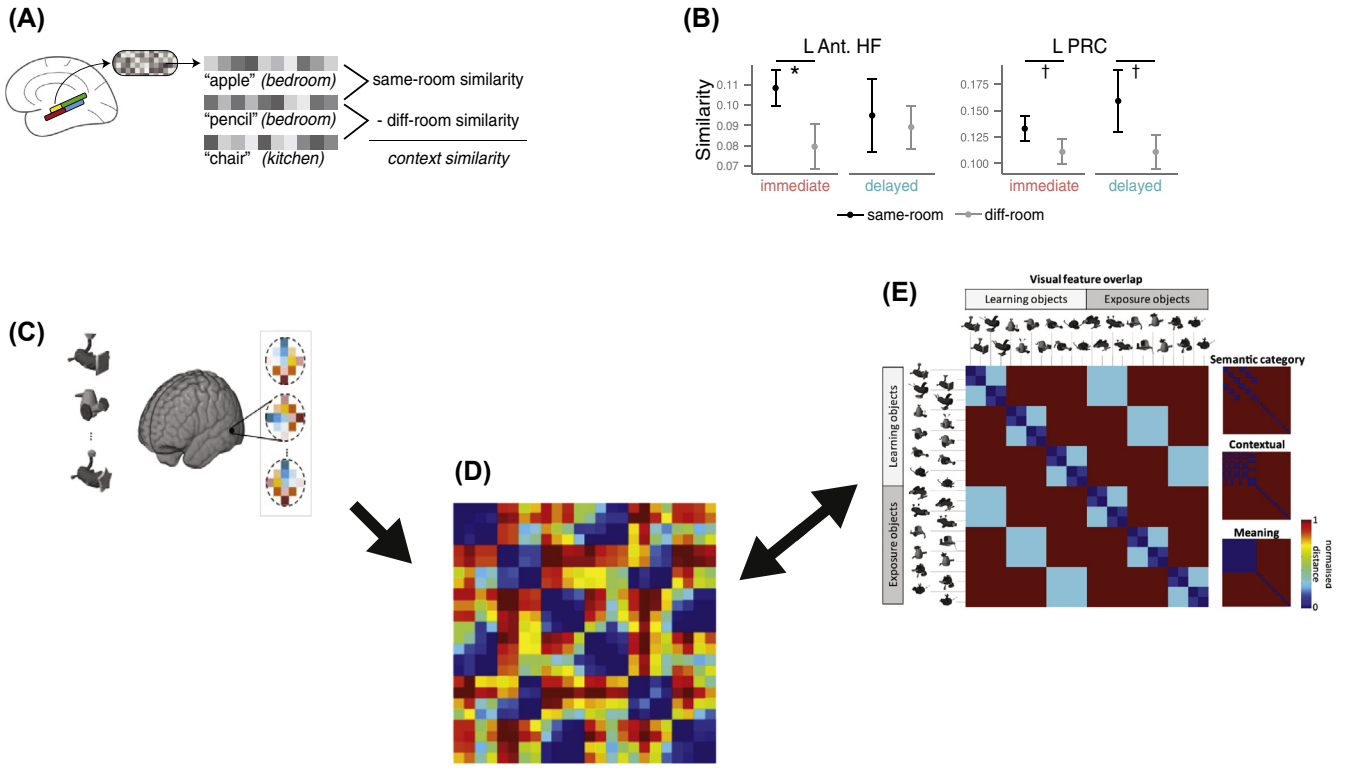

FIGURE 27.5 Different ways to test hypotheses with representational similarity analysis. (A) Extraction of patterns from a region of interest of interest for words studied in different room contexts. (B) Summarized pattern similarity (PS) values in left anterior hippocampus (L. Ant. HF) and left perirhinal cortex (L PRC) subdivided based on when an item was tested (immediate/delayed) and whether or not it occurred in the same or different room context. (C) Extraction of patterns using a searchlight for studied objects. (D) Resultant PS matrix for all searchlights. (E) Predicted matrix [or representational dissimilarity matrix (RDM)] that would be correlated with the observed patterns. Smaller matrices on the right show different potential models that could fit the data. *(A-B) Figure adapted from Ritchey et al. (2015) with permission of author. (C-E) Figure adapted from Clarke et al. (2016) with permission of author.*

For simpler experimental questions, you can directly extract the similarity values for pairs of trials within different conditions, average across them, and compare the mean level of PS between conditions [see Fig. 27.5A and B for an example (Ritchey et al., 2015)]. Note that if Pearson's r is the chosen similarity metric, the correlations must be z-transformed (Fisher transform) before they can be submitted to a parametric statistical test. This is because Pearson's coefficients are bounded between −1 and +1 and therefore are not normally distributed.

In general, you can conduct group-level analyses of PS values in the same manner as traditional univariate fMRI analyses. Typically, group analyses focus on a summary statistic approach in which a single mean contrast value is computed for each participant and entered into a group t-test. A preferable alternative is to take a mixed model analysis approach. In a mixed model, each trial pair within the condition is entered into the model as a single PS value, rather than averaged across a condition. The model then separately estimates variance attributable to the variable of interest, variance attributable to intersubject variability, and intertrial variance within each subject's data set. In general, these models are better suited to differentiate between interesting and uninteresting sources of variance in the data. A detailed treatment of mixed effects models is beyond the scope of this chapter, but interested readers can choose from several excellent papers on this topic (Barr et al., 2013; Bates, 2010; Bates et al., 2014; Singmann et al., 2015).

When using a searchlight analysis or examining data across several ROIs, it is important to correct for multiple comparisons. The most straightforward approach is to use nonparametric approaches such as permutation tests. These approaches often involve some sort of shuffling procedure to create a null distribution that is specific to the range of variance in the data and then the observed statistic being evaluated for significance can be compared to this generated distribution. Others have written more extensively about the nuances of computing and reporting permutations and other nonparametric statistics [(Etzel and Braver, 2013; Etzel et al., 2009) for an example application of this, see (Huffman and Stark, 2017)].

3. PITFALLS, PROBLEMS, AND PROPOSED SOLUTIONS

3.1 Noise Reduction

Although it is not a widely recognized issue in RSA, we feel it is important to consider the fact that fMRI data are inherently noisy. Even a large perceptual or cognitive signal will, at best, elicit an observed BOLD signal change around 2%−4% (Pillai and Zaca, 2011). In typical memory studies, signal changes are typically around 0.25%−0.5%. Therefore, it can be important to take some additional steps to remove sources of noise and nuisance variance from the data. Although our laboratory routinely employs many of these techniques, it is also our experience that these steps may have small effects. We have never seen noise reduction methods introduce an effect where one did not exist before, and we would be concerned if they did! We have seen, however, that careful data cleaning can enhance signal-to-noise ratios and sensitivity to relatively small signals in task- and pattern-based fMRI.

3.2 Voxel-Level Cleaning

One cleaning step is to identify and remove noisy voxels that consistently have fluctuations that exceed some range of the data (e.g., greater than 3 standard deviation mean signal change across all runs). Such large fluctuations are usually caused by motion artifacts. For example, a voxel near the edge of the brain or near a zone of signal dropout might alternate between sampling the brain and sampling nonbrain regions over the course of scanning. One way to identify these voxels is to look at each voxel's time course. If there are truly voxels that are contributing noise, their mean signal (or mean standard deviation in signal) should lie outside of the distribution of all other voxels. However, it is important to be sure that this variability is not correlated with your task (e.g., motion that co-occurs with a particular trial type). Typically, our laboratory has used cut-offs of 3−5 standard deviations outside of the range of the rest of the data. In our experience, one rarely needs to exclude voxels, and large fluctuations in many voxels are often indicative of a deeper issue (e.g., excessive movement artifact, misalignment of ROIs to functional data).

Instead of excluding voxels that are most likely contributing noise, an alternative approach is to only use the voxels that are most likely to be contributing signal. This approach, called feature selection, is relatively common in classification-based multivariate approaches (Norman et al., 2006), especially when the entire brain is being used for classification. Feature selection is not commonly used in RSA, and, in our limited tests, we have not found it to make a significant difference.

If you use feature selection, it is important to select voxels in an unbiased manner (Kriegeskorte et al., 2009). For example, if you scanned a separate run of another memory task where you could identify memory-sensitive voxels unrelated to memory for items in the task at hand, or if you had a separate localizer run to identify voxels that were selective to another

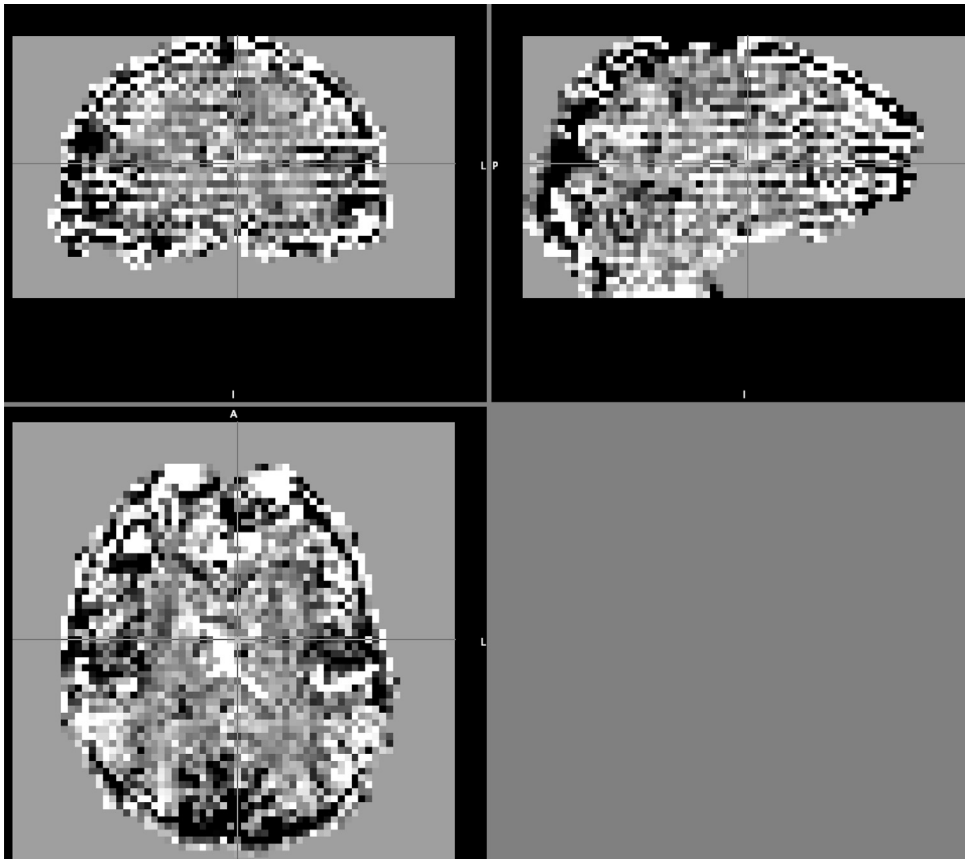

FIGURE 27.6 Example of a "stripey" beta that was identified as an outlier and removed from analysis pipeline. Visualized in FSL (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki). *Unpublished data, courtesy of Marika C. Inhoff.*

manipulation in your task, using these separate data to identify voxels would not be problematic. However, it is rare that we have the luxury of extra time in the scanner to collect these additional data.

3.3 Trial-Level Cleaning

Scaling up from the level of voxels, we can also think about noise at the trial level. When you separately model single-trial activation values, there will always be some trials that are poorly fit by the model, or trials that, for whatever reason, have extreme activation estimates. Our approach in removing outlier trials is identical, in principle, to removing outlier voxels. That is, to identify an outlier trial, we look at the distribution of beta values (e.g., by computing the mean beta value for each trial within a ROI) both for each subject, and, across the entire sample of subjects, to determine a threshold cut-off. We have also tried using a more standardized approach to determine a cut-off based on a z-score approach or percentile rank. What is important is that you pick a metric that is sensible for your data and be consistent in its application. If you are removing more than about 10%–15% of the betas for each participant, this typically indicates that there are other issues with the data. It is also a good idea to visually inspect the betas that are identified for removal. For example, our laboratory detected a scanner artifact (see Fig. 27.6) that resulted in "stripey" patterns in the modeled betas. These betas were clearly nonneural sources of noise and should be eliminated from the rest of the analysis pipeline.

Another approach to reduce the influence of noisy voxels is to use t-maps rather than beta maps for extracting patterns of interest (Walther et al., 2015). Whereas feature selection removes entire voxels from the analysis, one can instead include all the voxels and use t-values to weight the analysis toward voxels or trials that are generally more informative. A t-statistic is simply a beta that is scaled by the variance. This means that the most extreme voxels are penalized more than stable voxels because they vary more from time point to time point. Again, this approach should not produce a pattern where one did not exist previously but should enhance sensitivity to detect task-dependent PS effects (see Fig. 27.7 for an example in ventral visual cortex).

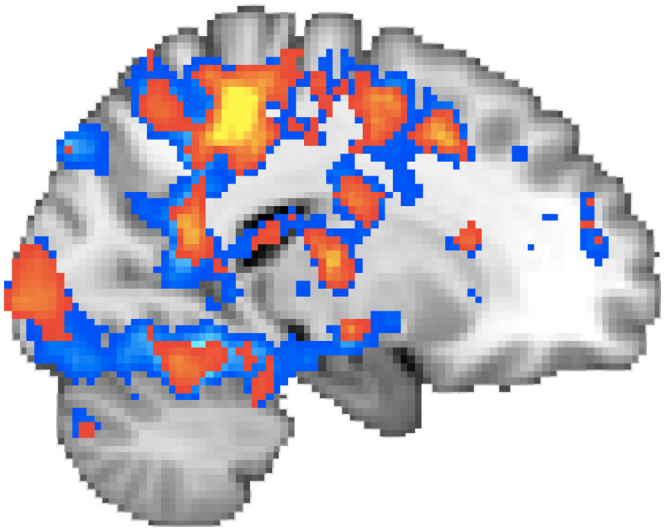

FIGURE 27.7 Informal comparison between representational similarity analysis (RSA) using betas (warm colors) versus t-maps (cool colors) comparing repeated presentations of the same item as compared with presentations of a different item. Notice that the RSA using t-maps does a better job of identifying PS effects in ventral visual cortex, as would be expected in this comparison. *Unpublished data, courtesy of Marika C. Inhoff.*

3.4 Unequal Trial Numbers Between Conditions

In memory paradigms, it is often necessary to compare pattern correlations to conditions that have unequal numbers of trial pairs. For instance, one could test whether hippocampal PS is higher for items that were associated with the same context than for trials that were associated with different contexts. Unfortunately, in this design, it would likely be the case that there are more trial pairs associated with different than the same context. A frequent concern in reviews, and in the general community, is whether an analysis might be biased to show higher correlations for the condition with fewer trial pairs than for the condition with more trial pairs. One way to address this concern is to rerun analyses in which you equate the number of trial pairs by subsampling conditions from the conditions with more trials (see Fig. 27.2 for an example). Even if the conditions vary in the number of trial pairs, results should remain stable if there are a reasonably large number of trial pairs in each condition.

3.5 Influence of Univariate Activity on Multivariate Patterns

In general, researchers expect univariate and multivariate fMRI analyses to reveal qualitatively different kinds of information (Arbuckle et al., 2018). A concern that is frequently expressed is whether a PS difference is confounded with a difference in the overall magnitude of activation between the two conditions. One might not necessarily want to separate the two influences, however. From a computational perspective, there is no reason to think that different representations in a neural network should be associated with equal overall firing rates. Nonetheless, many researchers believe that RSA can only be interpreted if pattern differences are not driven by overall activation magnitude.

As noted earlier, Pearson's r, Spearman's r, and cosine distance are magnitude insensitive. For example, if condition B shows the same pattern but more activation than condition A (i.e., the angle is preserved but the vector length differs), then the r value will be +1, just as if the two conditions were associated with the same activation values. That said, even when using metrics that are mathematically insensitive to activation magnitude, activation differences could nonetheless confound voxel PS estimates.

This is particularly the case when one is using a searchlight or an inappropriate ROI. For example, imagine a searchlight analysis in which the searchlight centers on a point between two anatomically distinct regions, A and B. Region A shows uniformly high activity in the memory condition but not in the control condition and Region B shows low activity in both conditions. RSA on the searchlight region, even using Pearson's r, would reveal a pattern difference across the memory and control conditions that are entirely driven by differences between overall (univariate) activation magnitude across the two regions spanned by the searchlight. Even in this case, the interpretation is not straightforward. One might conclude that a univariate effect in Region A is driving a spurious pattern difference between the memory and control conditions. Alternatively, one might conclude that there is an underlying representation in the memory condition that is distributed across a single brain region or network that is comprised of both Region A and Region B. Both conclusions are equally valid, depending on one's perspective.

From a practical standpoint, the issue of univariate activation confounds is often irrelevant. In some cases, the trial pairs are simply the recombination of all possible trials from one side of a univariate contrast (e.g., rearranging all remembered items based on their encoding condition, e.g., Dimsdale-Zucker et al., 2018). Of course, if there was a univariate difference between two different encoding conditions (e.g., had we found large activation differences between videos 1 and 3 and all trial pairs of interest came from these two videos), then there could be a univariate confound to be concerned about; however, with many conditions it is unlikely that this confound would arise by chance. In other cases, researchers have attempted to pull apart the univariate and multivariate contributions either by

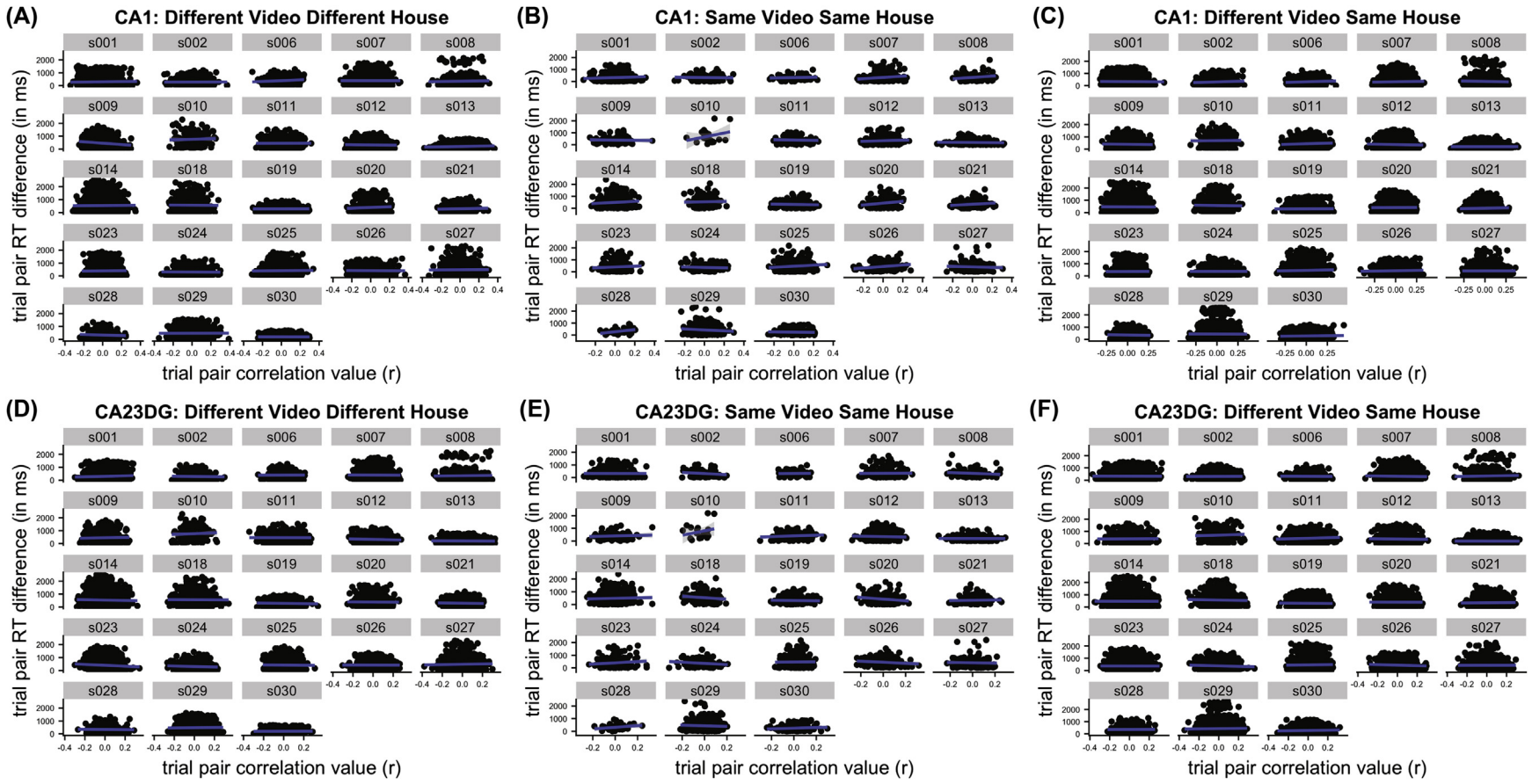

FIGURE 27.8 Correlations between differences in reaction time between pattern similarity (PS) trial pairs and PS values with linear fit lines (blue) in CA1(A-C) and CA23DG (D-F). *RT*, reaction time. *Figure adapted from Dimsdale-Zucker et al. (2018) with permission of author.*

removing the mean condition-wise effect (Macevoy and Epstein, 2009; Misaki et al., 2010) or showing that the information contained in the univariate and multivariate effects differ (Aly and Turk-Browne, 2016).

3.6 Influence of Subject-Specific Confounds

In 2013, Todd et al. (2013) showed that trial-wise reaction time differences (RT) could account for pattern differences between conditions, even if, on average, the two conditions do not show an RT difference. This seemingly counterintuitive finding reflects the fact that pattern classifiers can be driven by both positive and negative activation differences. Imagine a case where one subject is 100 ms faster in Condition A than B, whereas another subject is 100 ms faster on Condition B than A. A brain area where activation is sensitive to RT could be used to classify A and B states for both subjects, despite the fact that, across subjects, the average RTs are equal across conditions. In this case, we would not know whether accurate classification between Conditions A and B reflects interesting differences in underlying representations or whether it is a result of a relatively uninteresting effect that is related to individual variations in response latency or task difficulty.

Todd et al. considered whether this is an issue for nonclassification multivariate techniques such as RSA. Their conclusion was that RSA should be less likely to suffer from these types of confounds than classification-based methods. In the paper, they simply suggest including reaction time in a linear regression model to account for between-condition reaction time differences. Fig. 27.8 shows an example data set (Dimsdale-Zucker et al., 2018) where we included a random effect of reaction time in our mixed models (which is an essentially equivalent approach to linear regression). As you can see, in this data set it is clear that whether or not we account for reaction time does not change the observed pattern of effects. However, this a reasonable and straightforward control analysis to routinely perform if there are reaction time differences between conditions of interest.

4. REPRESENTATIONAL SIMILARITY ANALYSIS: WHERE DO WE STAND?

Since its introduction to the human imaging world (Kriegeskorte et al., 2006, 2008a), the number of publications using RSA or other multivariate techniques has steadily increased. It has been particularly useful in moving the field forward from localization of processes within regions to understanding how these regions represent information. Although we still know little about how neural activity drives large-scale differences in multivoxel activity patterns, RSA is a valuable tool for bridging the gap between human fMRI and invasive recording studies in animal models (Kriegeskorte et al., 2008b). RSA is now also beginning to be applied to single-unit recording studies of memory (McKenzie et al., 2016) and scalp EEG (Kaneshiro et al., 2015) showing that is a highly flexible analysis technique. In the future, we hope to see a continued use of RSA and other multivariate techniques to enhance our understanding of the mind and brain.

References

Abdulrahman, H., Henson, R.N., 2016. Effect of trial-to-trial variability on optimal event-related fMRI design: implications for Beta-series correlation and multi-voxel pattern analysis. NeuroImage 125, 756–766. https://doi.org/10.1016/j.neuroimage.2015.11.009.

Aguirre, G.K., 2007. Continuous carry-over designs for fMRI. NeuroImage 35 (4), 1480–1494. https://doi.org/10.1016/j.neuroimage.2007.02.005.

Aly, M., Turk-Browne, N.B., 2016. Attention stabilizes representations in the human hippocampus. Cerebr. Cortex 26 (2), 783–796. https://doi.org/10.1093/cercor/bhv041.

Arbuckle, S.A., Yokoi, A., Pruszynski, J.A., Diedrichsen, J., 2018. Stability of representational geometry across a wide range of fMRI activity levels. BioRxiv 266585. https://doi.org/10.1101/266585.

Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. J. Mem. Lang. 68 (3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bates, D., 2010. lme4: Mixed-Effects Modeling With R. Springer, Madison, WI.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2014. Fitting linear mixed-effects models using lme4. J. Stat. Software 1–48. https://doi.org/10.18637/Jss.V067.I01.

Blamire, A.M., Ogawa, S., Ugurbil, K., Rothman, D., McCarthy, G., Ellermann, J.M., et al., 1992. Dynamic mapping of the human visual cortex by high-speed magnetic resonance imaging. Proc. Natl. Acad. Sci. U. S. A. 89 (22), 11069–11073.

Bobadilla Suarez, S., Love, B.C., November 2017. Measures of neural similarity. In: Presented at the Society for Neuroscience, Washington, D.C. Retrieved from: http://www.abstractsonline.com/pp8/#!/4376/presentation/18122.

Burock, M.A., Buckner, R.L., Woldorff, M.G., Rosen, B.R., Dale, A.M., 1998. Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. Neuroreport 9 (16), 3735–3739.

Calhoun, V.D., Stevens, M.C., Pearlson, G.D., Kiehl, K.A., 2004. fMRI analysis with the general linear model: removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. NeuroImage 22 (1), 252–257. https://doi.org/10.1016/j.neuroimage.2003.12.029.

Clark, V.P., Maisog, J.M., Haxby, J.V., 1998. fMRI study of face perception and memory using random stimulus sequences. J. Neurophysiol. 79 (6), 3257–3265. https://doi.org/10.1152/jn.1998.79.6.3257.

Clarke, A., Pell, P.J., Ranganath, C., Tyler, L.K., 2016. Learning warps object representations in the ventral temporal cortex. J. Cognit. Neurosci. 28 (7), 1010–1023. https://doi.org/10.1162/jocn_a_00951.

Coutanche, M.N., Thompson-Schill, S.L., 2012. The advantage of brief fMRI acquisition runs for multi-voxel pattern detection across runs. NeuroImage 61 (4), 1113–1119. https://doi.org/10.1016/j.neuroimage.2012.03.076.

Dale, A.M., Buckner, R.L., 1997. Selective averaging of rapidly presented individual trials using fMRI. Hum. Brain Mapp. 5 (5), 329–340. https://doi.org/10.1002/(SICI)1097-0193(1997)5:5<329::AID-HBM1>3.0.CO;2-5.

Davatzikos, C., Ruparel, K., Fan, Y., Shen, D.G., Acharyya, M., Loughead, J.W., et al., 2005. Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. NeuroImage 28 (3), 663–668. https://doi.org/10.1016/j.neuroimage.2005.08.009.

Diana, R.A., Yonelinas, A.P., Ranganath, C., 2007. Imaging recollection and familiarity in the medial temporal lobe: a three-component model. Trends Cognit. Sci. 11 (9), 379–386. https://doi.org/10.1016/j.tics.2007.08.001.

Dimsdale-Zucker, H.R., Ritchey, M., Ekstrom, A.D., Yonelinas, A.P., Ranganath, C., 2018. CA1 and CA3 differentially support spontaneous retrieval of episodic contexts within human hippocampal subfields. Nat. Commun. 9 (1), 294. https://doi.org/10.1038/s41467-017-02752-1.

Edelman, S., Grill-Spector, K., Kushnir, T., Malach, R., 1998. Toward direct visualization of the internal shape representation space by fMRI. Psychobiology 26 (4), 309–321. https://doi.org/10.3758/BF03330618.

Etzel, J.A., Braver, T.S., 2013. MVPA permutation schemes: permutation testing in the land of cross-validation. In: Presented at the 3rd International Workshop on Pattern Recognition in Neuroimaging, Philadelphia, PA. https://doi.org/10.1109/PRNI.2013.44.

Etzel, J.A., Gazzola, V., Keysers, C., 2009. An introduction to anatomical ROI-based fMRI classification analysis. Brain Res. 1282 (Suppl. C), 114–125. https://doi.org/10.1016/j.brainres.2009.05.090.

Etzel, J.A., Zacks, J.M., Braver, T.S., 2013. Searchlight analysis: promise, pitfalls, and potential. NeuroImage 78, 261–269. https://doi.org/10.1016/j.neuroimage.2013.03.041.

Friston, K.J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M.D., Turner, R., 1998. Event-related fMRI: characterizing differential responses. NeuroImage 7 (1), 30–40. https://doi.org/10.1006/nimg.1997.0306.

Georgopoulos, A.P., Schwartz, A.B., Kettner, R.E., 1986. Neuronal population coding of movement direction. Science 233 (4771), 1416–1419.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293 (5539), 2425–2430. https://doi.org/10.1126/science.1063736.

Haynes, J.-D., Rees, G., 2005. Predicting the stream of consciousness from activity in human visual cortex. Curr. Biol. 15 (14), 1301–1307. https://doi.org/10.1016/j.cub.2005.06.026.

Heeger, D.J., Ress, D., 2002. What does fMRI tell us about neuronal activity? Nat. Rev. Neurosci. 3 (2), 142–151. https://doi.org/10.1038/nrn730.

Huffman, D.J., Stark, C.E.L., 2017. The influence of low-level stimulus features on the representation of contexts, items, and their mnemonic associations. NeuroImage 155 (Suppl. C), 513–529. https://doi.org/10.1016/j.neuroimage.2017.04.019.

Jonker, T.R., Dimsdale-Zucker, H.R., Ritchey, M., Clarke, A., Ranganath, C., 2018. Neural reactivation in parietal cortex enhances memory for episodically-linked information (under review).

Kaneshiro, B., Perreau Guimaraes, M., Kim, H.-S., Norcia, A.M., Suppes, P., 2015. A representational similarity analysis of the dynamics of object processing using single-trial EEG classification. PLoS One 10 (8), e0135697. https://doi.org/10.1371/journal.pone.0135697.

Kiani, R., Esteky, H., Mirpour, K., Tanaka, K., 2007. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. J. Neurophysiol. 97 (6), 4296–4309. https://doi.org/10.1152/jn.00024.2007.

Kriegeskorte, N., Cusack, R., Bandettini, P., 2010. How does an fMRI voxel sample the neuronal activity pattern: compact-kernel or complex spatiotemporal filter? NeuroImage 49 (3), 1965–1976. https://doi.org/10.1016/j.neuroimage.2009.09.059.

Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. U. S. A. 103 (10), 3863–3868. https://doi.org/10.1073/pnas.0600244103.

Kriegeskorte, N., Mur, M., Bandettini, P., 2008a. Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2 https://doi.org/10.3389/neuro.06.004.2008.

Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., et al., 2008b. Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60 (6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043.

Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S.F., Baker, C.I., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci. 12 (5), 535–540. https://doi.org/10.1038/nn.2303.

Leutgeb, S., Leutgeb, J.K., Moser, M.-B., Moser, E.I., 2005. Place cells, spatial maps and the population code for memory. Curr. Opin. Neurobiol. 15 (6), 738–746. https://doi.org/10.1016/j.conb.2005.10.002.

Libby, L.A., Hannula, D.E., Ranganath, C., 2014. Medial temporal lobe coding of item and spatial information during relational binding in working memory. J. Neurosci. 34 (43), 14233–14242. https://doi.org/10.1523/JNEUROSCI.0655-14.2014.

Macevoy, S.P., Epstein, R.A., 2009. Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. Curr. Biol. 19 (11), 943–947. https://doi.org/10.1016/j.cub.2009.04.020.

McKenzie, S., Keene, C., Farovik, A., Blandon, J., Place, R., Komorowski, R., Eichenbaum, H., 2016. Representation of memories in the cortical–hippocampal system: results from the application of population similarity analyses. Neurobiol. Learn. Mem. https://doi.org/10.1016/j.nlm.2015.12.008.

Misaki, M., Kim, Y., Bandettini, P.A., Kriegeskorte, N., 2010. Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. NeuroImage 53 (1), 103–118. https://doi.org/10.1016/j.neuroimage.2010.05.051.

Mumford, J.A., Davis, T., Poldrack, R.A., 2014. The impact of study design on pattern estimation for single-trial multivariate pattern analysis. NeuroImage. https://doi.org/10.1016/j.neuroimage.2014.09.026.

Mumford, J.A., Poline, J.-B., Poldrack, R.A., 2015. Orthogonalization of regressors in fMRI models. PLoS One 10 (4), e0126255. https://doi.org/10.1371/journal.pone.0126255.

Mumford, J.A., Turner, B.O., Ashby, F.G., Poldrack, R.A., 2012. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. NeuroImage 59 (3), 2636–2643. https://doi.org/10.1016/j.neuroimage.2011.08.076.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., Kriegeskorte, N., 2014. A toolbox for representational similarity analysis. PLoS Comput. Biol. 10 (4), e1003553. https://doi.org/10.1371/journal.pcbi.1003553.

Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cognit. Sci. 10 (9), 424–430. https://doi.org/10.1016/j.tics.2006.07.005.

Op de Beeck, H.P., 2010. Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? NeuroImage 49 (3), 1943–1948. https://doi.org/10.1016/j.neuroimage.2009.02.047.

Pillai, J.J., Zaca, D., 2011. DTI and fMRI: review of complementary techniques. In: Faro, S.H., Mohamed, F.B., Law, M., Ulmer, J.L. (Eds.), Functional Neuroradiology: Principles and Clinical Applications. Springer, New York, pp. 803–830.

Polyn, S.M., Natu, V.S., Cohen, J.D., Norman, K.A., 2005. Category-specific cortical activity precedes retrieval during memory search. Science 310 (5756), 1963–1966. https://doi.org/10.1126/science.1117645.

Rhodes, G., Byatt, G., Michie, P.T., Puce, A., 2004. Is the fusiform face area specialized for faces, individuation, or expert individuation? J. Cognit. Neurosci. 16 (2), 189–203. https://doi.org/10.1162/089892904322984508.

Ritchey, M., Montchal, M.E., Yonelinas, A.P., Ranganath, C., 2015. Delay-dependent contributions of medial temporal lobe regions to episodic memory retrieval. ELife 4, e05025. https://doi.org/10.7554/eLife.05025.

Singmann, H., Bolker, B., Westfall, J., 2015. afex: Analysis of Factorial Experiments [R Package].

Stark, S.M., Reagh, Z.M., Yassa, M.A., Stark, C.E.L., 2017. What's in a context? cautions, limitations, and potential paths forward. Neurosci. Lett. https://doi.org/10.1016/j.neulet.2017.05.022.

Todd, M.T., Nystrom, L.E., Cohen, J.D., 2013. Confounds in multivariate pattern analysis: theory and rule representation case study. NeuroImage 77, 157–165. https://doi.org/10.1016/j.neuroimage.2013.03.039.

Visser, R.M., de Haan, M.I.C., Beemsterboer, T., Haver, P., Kindt, M., Scholte, H.S., 2016. Quantifying learning-dependent changes in the brain: single-trial multivoxel pattern analysis requires slow event-related fMRI. Psychophysiology 53 (8), 1117–1127. https://doi.org/10.1111/psyp.12665.

Viswanathan, S., Cieslak, M., Grafton, S.T., 2012. On the geometric structure of fMRI searchlight-based information maps. Quantitative Biology. Retrieved from: http://arxiv.org/abs/1210.6317.

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., Diedrichsen, J., 2015. Reliability of dissimilarity measures for multi-voxel pattern analysis. NeuroImage. https://doi.org/10.1016/j.neuroimage.2015.12.012.

Worsley, K.J., Taylor, J.E., 2006. Detecting fMRI activation allowing for unknown latency of the hemodynamic response. NeuroImage 29 (2), 649–654. https://doi.org/10.1016/j.neuroimage.2005.07.032.

Yonelinas, A.P., 2002. The nature of recollection and familiarity: a review of 30 years of research. J. Mem. Lang. 46 (3), 441–517. https://doi.org/10.1006/jmla.2002.2864.

Zeithamova, D., de Araujo Sanchez, M.-A., Adke, A., 2017. Trial timing and pattern-information analyses of fMRI data. NeuroImage. https://doi.org/10.1016/j.neuroimage.2017.04.025.