NeuroImage 153 (2017) 221–231

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

# Trial timing and pattern-information analyses of fMRI data

Dagmar Zeithamova*, Maria-Alejandra de Araujo Sanchez, Anisha Adke

*University of Oregon, Department of Psychology, 1227 University of Oregon, Eugene, OR 97403, USA*

ARTICLE INFO

*Keywords:*
fMRI
Pattern classification
Decoding
Experimental design optimization
Multivoxel pattern analysis
Pattern similarity analysis
Representational similarity analysis
Memory
Object representation

ABSTRACT

Pattern-information approaches to fMRI data analysis are becoming increasingly popular but few studies to date have investigated experimental design optimization for these analyses. Here, we tested several designs that varied in the number of trials and trial timing within fixed duration scans while participants encoded images of animals and tools. Trial timing conditions with fixed onset-to-onset timing ranged from slow 12-s trials with two repetitions of each item to quick 6-s trials with four repetitions per item. We also tested a jittered version of the quick design with 4–8 s trials. We assessed the effect of trial timing on three dependent measures: category-level (animals vs. tools) decoding accuracy using a multivoxel pattern analysis, item-level (e.g., cat vs. dog vs. lion) information estimates using pattern similarity analysis, and memory effects comparing pattern similarity scores across repetitions of individual items subsequently remembered vs. forgotten. For single trial estimates, category decoding was equal across all trial timing conditions while item-level information and memory effects were better detected using slow trial timing. When modeling events on an item-by-item basis across all repetitions of a given item, a larger number of quick, regularly spaced trials provided an advantage over fewer slow trials for category decoding while item-level information was comparable across conditions. Jittered and non-jittered versions of the quick trial timing did not differ significantly in any analysis. These results will help inform experimental design choices in future studies planning to employ pattern-information analyses and demonstrate that design optimization guidelines developed for univariate analyses of a few conditions are not necessarily optimal for pattern-information analyses and condition-rich designs.

## Introduction

The possibility of non-invasive imaging of activity in a healthy human brain using functional MRI (fMRI) has transformed the field of psychology and neuroscience in recent decades. For most of this time, studies relying on traditional univariate analysis using the general linear model (GLM) framework have been fruitful in converging with knowledge gathered from animal-lesion and human-neuropsychological studies as well as generating new insights unique to observing the intact human brain. While the utility of the univariate approach has not been exhausted, new analytical approaches have emerged that consider multivariate patterns of activation across many voxels or entire regions. Multivoxel pattern analyses, or MVPA, can often provide a richer picture than univariate analyses and open the door for an array of new research questions (Haxby et al., 2001; Jimura and Poldrack, 2012; Kriegeskorte and Bandettini, 2007; Norman et al., 2006; Ward et al., 2013).

The two common types of MVPA are classifier-based approaches and pattern-similarity approaches. Classifier-based approaches use machine-learning pattern-classification algorithms, such as support vector machines or logistic regression classifiers, to decode information from distributed activation patterns (Norman et al., 2006; Pereira et al., 2009). A machine-learning algorithm is trained to differentiate between patterns of activation across voxels evoked by two or more conditions (e.g., viewing of faces vs. viewing of scenes). The analysis is typically done at a single subject level by training the classifier on data

*Abbreviations:* AC, Anterior Commissure; ANTs, Advanced Neuroimaging Tools; BET, Brain Extraction Tool; BOLD, Blood-Oxygen-Level Dependent; df, degree of freedom; FEAT, fMRI Expert Analysis Tool; fMRI, Functional Magnetic Resonance Imaging; FOV, Field Of View; FSL, FMRIB Software Library; GG, Greenhouse-Geisser corrected for non-sphericity; GLM, General Linear Model; GRAPPA, Generalized Autocalibrating Partial Parallel Acquisition; hip, Hippocampus; ifg, Inferior frontal gyrus; ipar, Inferior parietal cortex; med (3@8 s), Experimental design where each item is presented 3 times, and trial length is 8 s; MPRAGE, Magnetization-Prepared Rapid Gradient-Echo; MVPA, Multivoxel Pattern Analysis; PC, Posterior Commissure; phc, Parahippocampal cortex; pfus, Posterior fusiform; PSA, Pattern Similarity Analysis; quick (4@4–8 s), Experimental design where each item is presented 4 times, and trial length is either 4 s or 8 s; quick (4@6 s), Experimental design where each item is presented 4 times, and trial length is 6 s; ROI, Region of Interest; RSA, Representational Similarity Analysis; SOA, Stimulus-Onset Asynchrony; slow (2@10 s), Experimental design where each item is presented 2 times, and trial length is 10 s; slow (2@12 s), Experimental design where each item is presented 2 times, and trial length is 12 s; supar, Superior parietal cortex; SVM, Support Vector Machines; TE, Echo Time; TI, Inversion Time; TR, Repetition Time; UO, University of Oregon

* Corresponding author.
*E-mail addresses:* dasa@uoregon.edu (D. Zeithamova), dearaujo@uoregon.edu (M.-A. de Araujo Sanchez), aadke@uoregon.edu (A. Adke).

http://dx.doi.org/10.1016/j.neuroimage.2017.04.025
Received 30 November 2016; Accepted 11 April 2017
Available online 12 April 2017
1053-8119/ 

from all but one run and then testing the generalization accuracy of the classification on the remaining run. This process is repeated for each subject until all runs have been left out once and used to test the classifier. A cross-validated accuracy is then computed by averaging the generalization accuracy across all iterations. Above-chance classification accuracy within a region demonstrates the sensitivity of that region to the tested conditions.

Pattern similarity analysis (PSA, also referred to as representational similarity analysis, or RSA) aims to determine what information is coded in a region by estimating the activation pattern associated with specific stimuli or conditions (neural item representations) and then computing similarity between these patterns (Kriegeskorte et al., 2008a). For instance, if an experiment consists of presenting a series of objects, it may be of interest not only how the brain responds to objects in general, but also how patterns of activation relate to specific objects (e.g., "banana" vs. "orange" vs. "basketball"). Determining which stimuli are treated as similar and which are treated as different can provide insight into which stimulus features various regions are most sensitive to. For example, if the activation pattern evoked by an orange is more similar to the pattern evoked by a basketball compared to the pattern evoked by a banana, the region may code for perceptual properties of the stimuli such as shape or color. However, if an orange-evoked activation pattern is more similar to the pattern evoked by a banana than to the pattern evoked by a basketball, then the region may code for conceptual information such as belonging to the category of fruits. This is just one of the examples that demonstrates the utility of these multivariate analyses to utilize the potential information embedded in fMRI datasets.

While several studies have been dedicated to fMRI design optimization, most have focused on optimization for standard univariate analyses (Burock et al., 1998; Dale, 1999; Friston et al., 1999; Stark and Squire, 2001). It is unclear whether designs optimized for univariate contrasts of a few experimental conditions are necessarily optimal for pattern-information analyses and vice-versa. For example, one study showed that greater numbers of shorter runs work better for classifier-based analyses than smaller numbers of longer runs (Coutanche and Thompson-Schill, 2012). This is presumably because a larger number of shorter runs provide more training trials for leave-one-run-out cross validation, thus resulting in more robust estimates of condition-associated patterns. Such consideration is irrelevant for standard univariate analyses because they typically estimate condition-associated activation using the entire dataset and rarely include cross-validation. In fact, fewer, longer runs may be better suited for standard univariate analyses (Henson, 2007). Accordingly, a plan to employ pattern-information analyses should be considered when designing an fMRI experiment, but there is currently limited empirical evidence to guide design choices.

There are many different approaches and applications of pattern-information analyses, making it impossible to devise universal recommendations. However, one aspect employed by many studies that is unique to pattern-information analyses is a condition-rich design. Commonly, standard univariate designs employ a small number of experimental conditions, each consisting of many trials. Trials are spaced within and between conditions so that a regressor representing one condition is minimally correlated with regressors representing other conditions and the contrast between conditions is detectable with a realistic effect size. In contrast, pattern-information analyses often consider each unique stimulus or each individual trial as a separate condition. For instance, rather than measuring the BOLD response to objects in general, patterns of activation related to "cat" vs. "dog" vs. "hammer" are of interest. Therefore, some of the established procedures for trial timing optimization for univariate analyses (Dale, 1999) may not transfer well to condition-rich designs.

One solution are event-related designs that include sufficient time between individual events so that the BOLD response can return close to baseline before the next event is presented (Buckner et al., 1996; McCarthy et al., 1997). Slow event-related designs, with longer onset-to-onset timing (stimulus-onset asynchrony, or SOA) should be appropriate for both univariate and multivariate approaches, as they allow for an estimation of trial-specific activation patterns without much contamination from neighboring trials. Prior research has shown that longer trials may be necessary to estimate single trial activation, especially when trial-by-trial changes are of interest (Visser et al., 2016). Slow event-related designs have a significant drawback because only a few stimuli can be presented in a fixed amount of time (about 3–5 stimuli per minute), leading to fewer items or fewer repetitions per item that fit within the experiment. In addition, slow trial pacing allows for more mind-wandering, which may affect performance (Antrobus et al., 1966) and be reflected in the BOLD signal in an uncontrolled way.

Another popular approach is to sacrifice individual-trial estimability and instead quickly present a large number of trials. When inter-stimulus time is jittered and conditions are appropriately intermixed to ensure that the predicted BOLD signals for different conditions are not strongly correlated, rapid event-related designs can provide better sensitivity to detect univariate differences between conditions (Burock et al., 1998; Dale, 1999). A modified version of this approach, coined as a "quick" event-related design, has been successfully used in seminal work using PSA to compare representations of individual objects across humans and monkeys (Kriegeskorte et al., 2008b). In this study, the researchers presented the same set of 92 images of real-world objects (e.g., ear, kiwi, white male face) to monkeys and humans while recording inferotemporal cortex responses using single-cell recording in monkeys and high-resolution fMRI in humans. Measuring the pairwise similarity of the neural patterns in response to each object, the authors demonstrated a hierarchy of categorical object representations in the inferotemporal cortex strongly matched between species. However, in this study, each individual stimulus was presented 11–14 times and modeled using a single regressor across repetitions. In that respect, the number of trials per regressor was not much smaller than what is typically used in univariate analyses. Thus it is unclear whether such trial timing could be used to estimate item-level representation on single trials.

Based on prior research (Abdulrahman and Henson, 2016; Visser et al., 2016), we hypothesized that if items have fewer repetitions or if the estimation of each individual trial is of interest, rapid/quick designs might not be suitable. Rapid event-related designs have several downsides. First, variable jitter may affect cognitive processes differentially for long vs. short trials. For instance, participants may not fully process the current trial or may not be ready for the next trial after short delays. Conversely, they may grow impatient awaiting the next trial during an unusually long delay. These trial-by-trial differences due to changes in SOA can affect behavioral performance (see Suppl. Fig. 1) and likely neural responses as well. While trial-specific effects can average out across multiple trials in the same condition, they would remain part of the activation patterns estimated based on a single (or a few) presentation(s). Finally, rapid presentations of individual items lead to a strong overlap of subsequent activation patterns due to sluggish BOLD responses, which may make trial-by-trial estimates less reliable. While iterative modeling of individual trials (LSS method, Mumford et al., 2012) may, to some degree, alleviate the challenge of response overlap, subtle trial-by-trial effects may still be lost (Visser et al., 2016) and the variable SOA effects would not be addressed.

Notably, neither jittered trial onsets nor long SOAs are used in behavioral cognitive experiments outside of an fMRI context. Instead, trials are presented regularly at a pace that is dictated by psychological constraints. Because of the shortcomings of these designs discussed above, we were interested in whether it would be possible to deviate from these common fMRI standards and use trial timings that are more convenient from a psychological perspective without compromising decoding efficiency. We reasoned that the main motive for including jitter is to ensure a unique predicted BOLD response for each

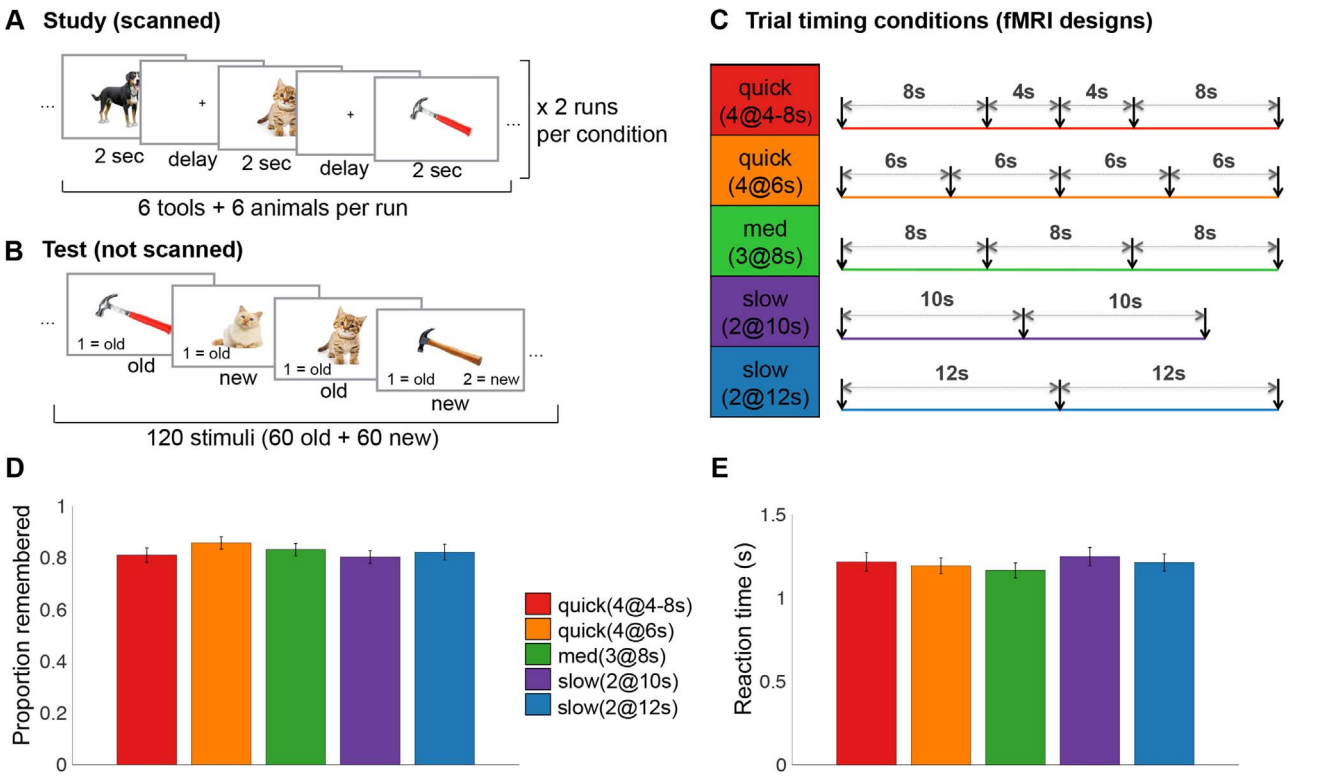


**Fig. 1. Behavioral methods and memory performance.** A. Encoding. During each run, participants were presented with 12 different items (6 animals and 6 tools) repeated 2–4 times. The number of repetitions of each item and the trial timing varied between runs, with two consecutive runs always using the same 12 items and the same timing condition in a different order. B. During the self-paced test, participants were presented with the 60 items they had studied and 60 similar lures randomly intermixed, and were asked to provide an old/new judgment. C. Trial timing conditions. Each trial timing condition included 2–4 repetitions per item. Fewer repetitions were presented using trials with a longer SOA, more repetitions were presented using trials with a shorter SOA. A schematic example of possible trial onsets (vertical arrows) within a 24-s window is shown next to each design label. Horizontal arrows show the trial timing (onset to onset). All designs except slow (2@10 s) fit within the same duration scan of 288 s. D. Proportion of items remembered for each design, averaged across participants. E. Response times by design, averaged across participants. In both D and E, error bars denote across-subject standard error of the mean.

condition, minimally correlated with other conditions. This consideration is less relevant for condition-rich designs. When the activation pattern related to each trial is of interest, and thus each trial is its own condition, jitter would presumably improve the estimability of one trial only at an expense of other trials that would be pushed closer together. When interested in patterns related to individual items repeated across several trials, random ordering of repetitions may serve the same purpose as jitter to ensure that predicted BOLD signals related to different items are uncorrelated. We thus hypothesized that trial jittering may not be necessary or beneficial for decoding analyses. Instead, it may be more appropriate to evenly split the null time across all trials, to make the shortest inter-stimulus intervals longer and to remove effects related to SOA variability. At the opposite end, slow trials are equally inconvenient from a psychological perspective and severely restrict the total number of trials that can fit into an experiment. We reasoned that when trials are spaced at least 6 s apart, a peak of one trial would presumably fall on (or before) the onset of the next trial, when the BOLD signal related to that trial would not yet be detectable. Thus, we wanted to test the possibility of obtaining reliable trial-by-trial pattern estimates using a fixed SOA starting at 6 s. Based on prior research (Abdulrahman and Henson, 2016; Burock et al., 1998; Visser et al., 2016), we did not test shorter fixed SOAs as they are likely unsuitable for both univariate effect detection and trial-by-trial pattern analyses.

The current study tested the empirical effects of trial timing ranging from quick to slow event-related, trading off the length of the SOA with the number of repetitions of individual stimuli. We tested the trial timing effects using a memory encoding task in which participants intentionally encoded multiple exemplars from two categories: animals (cat, dog, lion, etc.) and tools (mallet, scissors, etc.). This task allowed us to evaluate the trial timing effects on both item- and category-representation estimates. To evaluate the reliability of the estimates and to avoid inflated Type I error associated with within-run decoding (Mumford et al., 2014), two runs of each design were completed using the same physical stimuli in a different order. Item and category estimates were then compared across runs. Category representations were indexed as the classifier-based cross-validation accuracy for decoding brain patterns related to animals vs. tools. Item-level representations were assessed using PSA and indexed as the difference between across-run neural pattern similarity for the same item (e.g., cat-cat similarity across runs) vs. neural pattern similarity of different items from the same category (e.g., cat-dog similarity across runs). Finally, to test the suitability of each design to detect more subtle effects on item-representation estimates, we tested memory effects on pattern similarity, indexed as differences in similarity scores for items that were subsequently remembered vs. forgotten (Xue et al., 2010).

## Materials and methods

### *Subjects*

Thirty-eight young, healthy adults (mean age 21.8, range 18–35, 19 females) from the University of Oregon (UO) and the surrounding community participated in the study. All participants signed an informed consent and were screened for normal or corrected-to-normal vision, no history of neurological disease, and no contra-indication for MRI in concordance with the Standard Operating Procedures of the UO Lewis Center for Neuroimaging. The study was approved by the UO Institutional Review Board. Two participants were excluded from the study because they fell asleep during at least one scan. One participant was excluded because their behavioral data were accidentally corrupted. Data from the remaining 35 participants are reported throughout the manuscript.

*Stimuli*

The stimulus set consisted of 120 colored images of 60 tools and 60 animals. Within the set, there were two versions of each tool or animal, which differed somewhat in their appearance (e.g., two different giraffes, two different mallets).

*Procedure*

Participants intentionally encoded one version of each tool and animal (30 tools+30 animals) during functional scanning (Fig. 1**A**). The stimuli were presented one at a time for 2 s, followed by a delay that corresponded to one of the five trial timing conditions. Participants were asked to try to remember the images as best they could because they would be later tested on them. Participants did not make any responses to the images during scanning. After the conclusion of the encoding phase, participants completed a self-paced old/new recognition test outside the scanner. The test included all 60 encoded images as well as 60 similar lures (e.g., a different cat, a different dog, a different mallet) to ensure that participants remembered the specific image (e.g., the specific cat image rather than just remembering having seen a cat; Fig. 1**B**). Target and lure images from all conditions were randomly intermixed. The test was used to back-sort the encoding items as remembered (target endorsed and lure rejected) or forgotten (target rejected and/or lure endorsed). Targets and lures were counterbalanced across subjects.

Our main interest was to compare, within-subjects, the utility of different fMRI designs for pattern information analyses. For this purpose, the 60 unique encoding items were randomly split into 5 sets (12 stimuli, 6 tools and 6 animals, in each set) for each participant. Each set was assigned to one of five different presentation schedules (fMRI designs), four of which had the same total length of 288 s per run and one of a shorter scan length of 240 s per run.

Across the five trial timing conditions, we systematically varied the number of item repetitions and the onset-to-onset spacing of trials (SOA) within a run (Fig. 1**C**). There were always 12 unique items repeated 2–4 times per run. To test the reliability of item and category estimates from each design, we ran the same design, with the same 12 items in a different order, in two consecutive runs. After that, a new set of 12 items was presented across the next two runs, using a different trial timing condition. This allowed us to evaluate pattern similarities without contamination by autocorrelation across time points within a single run and to avoid potential inflation of Type I error when decoding is done within a single run (Mumford et al., 2014). The order of the designs was counterbalanced across subjects so each design appeared with equal frequency as the first, second, third, fourth or fifth design during encoding.

The timing varied from slow designs with 2 repetitions per item (12-s SOA in slow (2@12 s), 10-s SOA in slow (2@10 s)), through a medium design with 3 repetitions per item and 8-s SOA (med (3@8 s)), to a quick design with 4 repetitions per item and 6-s SOA (quick (4@6 s). Slow (2@10 s) is not fully comparable to the other trial timing conditions as any differences in that design could be attributed to shorter total scan duration (240 s vs. 288 s for all other trial timing conditions). However, we included it in the main report for ease of comparison as its inclusion or exclusion in the statistical analyses did not change any of the conclusions.

While testing the effect of jitter across the full range of SOA values would be beyond the scope of the current study, we did include a jittered quick design. We modeled the timing in this design after a seminal study on pattern similarity analysis of categorical representation in visual cortices (Kriegeskorte et al., 2008b). It included a short bookend null time at the beginning and end of each run and several 4-s fixation trials interspersed between 4-s task trials (2 s stimulus+2 s fixation). We also borrowed the term "quick" design from that study, as the 2 s stimulus presentations could be in principle even closer in time than 4 s in a true "rapid" design. As a result, trial onsets in this trial timing condition were spaced 4 or 8 s apart, denoted as quick (4@4–8 s). Jittered and non-jittered quick designs had the same number of trials and the same total scan length, the only difference being that the null time (including null trials and bookend time) in the jittered design was distributed evenly across trials in the non-jittered version. Within each analysis (detailed below), we also tested pair-wise comparisons between conditions, including the comparison of jitter vs. non-jittered quick design to assess the effect of jitter in designs with a quick SOA.

*MRI scanning*

Scanning was completed on a 3T Siemens Skyra at the UO Lewis Center for Neuroimaging using a 32 channel head coil. Head motion was minimized using foam padding. The scanning session started with a localizer scan followed by a standard high-resolution T1-weighted MPRAGE anatomical image (TR, 2500 ms; TE, 30 ms; TI, 1100 ms, flip angle, 7°; matrix size, 256×256; 176 contiguous sagittal slices; FOV, 256 mm; slice thickness, 1 mm; voxel size 1.0×1.0×1.0 mm; GRAPPA factor, 2). Functional data were acquired during the encoding portion of the task using a multiband gradient-echo pulse sequence (TR, 2000 ms; TE, 26 ms; flip angle, 90°; matrix size, 100×100; 72 contiguous slices oriented 15° off the AC-PC line to reduce prefrontal signal dropout; interleaved acquisition; FOV, 200 mm; voxel size: 2.0×2.0×2.0 mm; GRAPPA factor, 2; Multi-band acceleration factor, 3). The visual stimuli were projected onto a screen that was viewed through a mirror. Each subject completed 10 runs (5 trial timing designs, 2 runs of each). Scanning concluded with a custom anatomical T2 coronal image (TR, 13520 ms; TE, 88 ms; flip angle, 150°; matrix size, 512×512; 65 contiguous slices oriented perpendicularly to the main axis of the hippocampus; interleaved acquisition; FOV, 220 mm; voxel size, 0.4×0.4×2.0 mm; GRAPPA factor, 2). The recognition test was completed outside the scanner.

*fMRI preprocessing and analysis*

Dicom files were converted to nifti format using the "dcm2nii" function from MRIcron (https://www.nitrc.org/projects/mricron). fMRI preprocessing and data analysis were carried out using FEAT (fMRI Expert Analysis Tool), version 6.00, part of FSL (www.fmrib.ox.ac.uk/fsl). The following steps were applied: motion correction within each run using McFlirt; non-brain removal using BET; grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor; highpass temporal filtering (sigma=100 s) and minimal spatial smoothing (sigma=1). The MPRAGE anatomical scan of each subject was coregistered to their first functional volume by rigid/affine transformations using the Advanced Neuroimaging Tools (ANTs: http://stnava.github.io/ANTs/). The preprocessed functional scans were used in first-level GLM models (described below) that included FILM pre-whitening with local autocorrelation-correction as well as motion parameters as regressors of no interest.

Two models were fit to each run, one to estimate the pattern of activation associated with each item (e.g., cat, dog, mallet) and one to estimate the pattern of activation evoked by each individual trial (i.e., trial 1, trial 2, etc). We were especially interested in the degree to which it is possible to decode category-level and item-level information on an individual trial basis. In order to estimate the pattern of activation evoked by each trial, we constructed a GLM that modeled each trial as a separate regressor (Fig. 2**A**). This model is also known as the least square-all (LSA) trial model as it includes all single-trial regressors in one model. This trial (LSA) model included 24 regressors in the Slow designs, 36 regressors in the Med design, and 48 regressors in the Quick designs. Each regressor was modeled as a stick function with a single "stick" of a unit length, convolved with the canonical hemodynamic response function as implemented in FSL (gamma function with a phase of 0 s, a standard deviation of 3 s and a mean lag of 6 s).

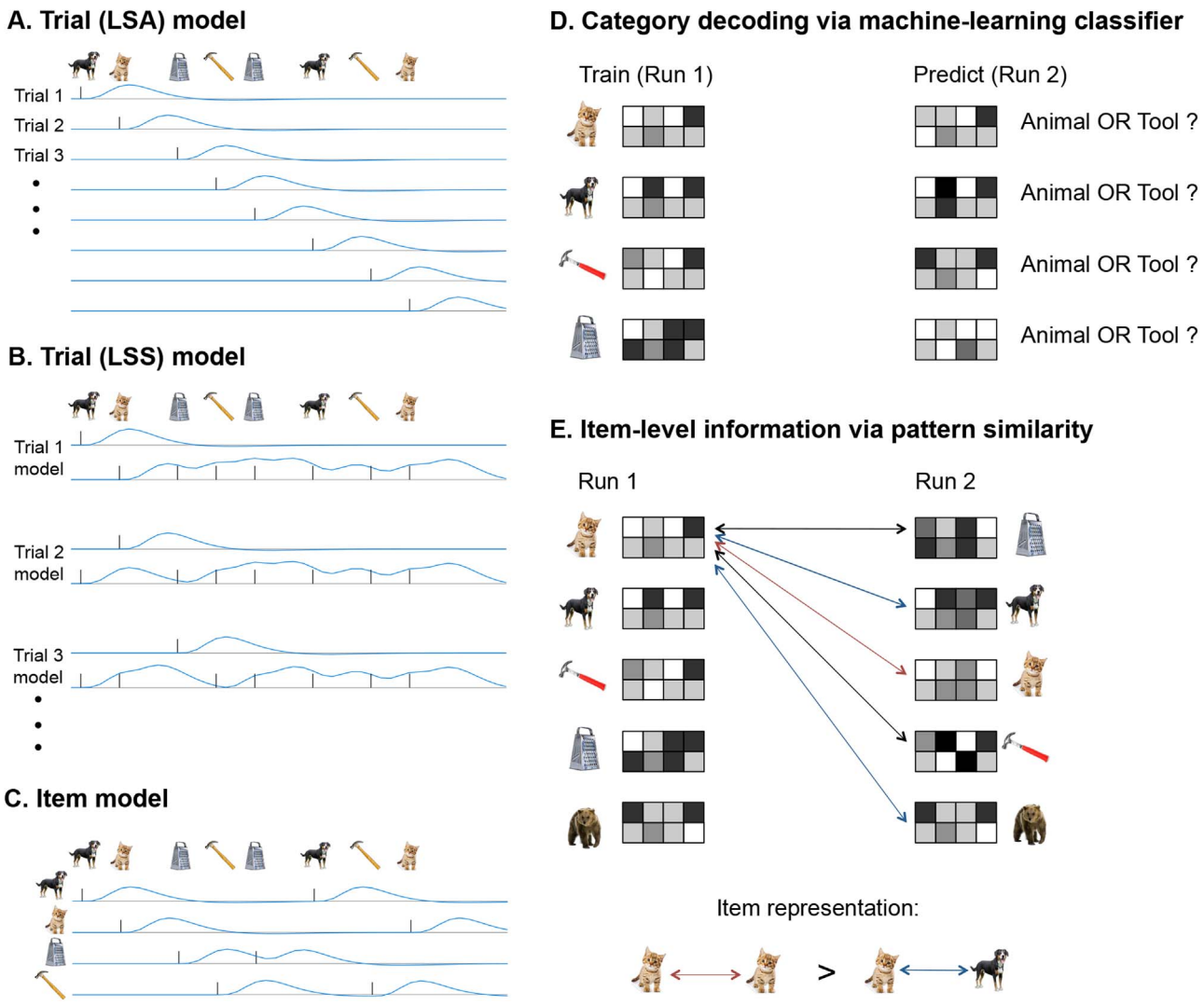


**Fig. 2. Analysis strategy. A**. Trial (LSA) model. BOLD response modeled using one regressor for each trial. **B.** Trial (LSS) model. A separate model is used to obtain an estimate for each trial. **C.** Item model. BOLD response modeled using one regressor for all repetitions of each unique item. **D.** Category decoding via machine learning. The classifier is trained on the activation patterns of one run and tested on the patterns from the other run of the same design. **E.** Indexing item-level information. Pairwise neural pattern similarities were computed for all trials (from the trial model) or all items (in the item model) across runs. Item representation indices were computed as the difference between the neural similarity of patterns evoked by the same item in separate runs (red arrow) and the neural similarity evoked by two different items from the same category (blue arrows). For example, if there is item-level information, the pattern of activation evoked by a cat from run 1 should be more similar to the pattern evoked by a cat from run 2 than to those evoked by other animals.

In deciding whether to include or not include temporal derivatives in the GLM model, we chose the model that provided the best fit within each design, in order to assess differences in decoding performance that may persist even when the best model for the design is used. We thus ran the trial model both with and without temporal derivatives and extracted the contrast values for task vs. baseline in the lateral occipital cortex for each trial timing condition as an empirical metric of model fit orthogonal to our results of interest. More robust visual activation was estimated using the model without temporal derivatives in quick (4–8 s) and quick (6 s) designs, and using a model with temporal derivatives in med (8 s), slow (10 s) and slow (12 s) designs. These effects were statistically significant for all but the quick (6 s) design, which showed very comparable results with either approach (Suppl. Fig. 2A). We thus included temporal derivatives only in the med (8 s), slow (10 s) and slow (12 s) designs. The empirical model fit estimate based on overall task-related activation ended up being consistent with our metrics of interest (category representation, item representation, memory effects), albeit the effect was much smaller. In other words, excluding temporal derivatives when modeling quick designs and including temporal derivatives when modeling slower designs yielded optimal decoding performance for the given design. This performance was then compared across conditions.

A prior study (Mumford et al., 2012) suggested that single-trial estimates may be more reliable when they are extracted iteratively, where a separate model is constructed and estimated for each trial. The sole trial of interest is modeled in one regressor and all other trials are modeled together as a second regressor (Fig. 2**B**). This modeling approach is also known as least square-single (LSS) because only a single trial is modeled as its own regressor during each iteration. Another study (Abdulrahman and Henson, 2016) showed that LSA vs. LSS modeling approaches may be advantageous under different conditions. Thus, we also generated trial beta-series using this iterative trial (LSS) model to test whether this approach may benefit some trial timing conditions or analyses specifically. Consistent with prior reports (Abdulrahman and Henson, 2016; Mumford et al., 2012), we found both advantages and disadvantages of the LSS model overall, but the pattern of results across conditions was similar to that obtained using the trial (LSA) model. We thus provide detailed LSA results and a summary of LSS results in the main manuscript and present the detailed LSS results in the Supplement. Consistent with prior approaches (Abdulrahman and Henson, 2016; Mumford et al., 2012), temporal derivatives were not included in LSS models.

Finally, our design also allowed us to compare how the trade-off

between the SOA and the number of repetitions of each item affects the estimability of neural activation patterns related to each item combined across repetitions. To estimate the activation related to each item, we constructed an item model (Fig. 2**C**) consisting of 12 regressors in all runs, one for each unique item encoded in that run (e.g., a cat regressor, a dog regressor, a mallet regressor, etc). While there were always 12 unique items encoded, the number of times each item was presented varied between 2 (for slow designs) and 4 (for "quick" designs). Thus, each regressor was modeled as a stick function with either 2, 3 or 4 "sticks" of a unit length, convolved with the same canonical hemodynamic response function as the trial (LSA) model. Temporal derivatives in the item model had similar effects as in the trial (LSA) model (Suppl. Fig. 2B). We thus included temporal derivatives only in the medium and slow designs. For completeness, we are providing results with and without TD for every condition for both trial and item models in the Supplement (Suppl. Fig. 7).

After estimating all models, we generated the beta-series representing each trial (in the trial model) or each item (in the item model) within a run. ANTs was used to estimate realignment parameters between an example functional image from each run and the example functional image from the first run of every subject. These parameters were then applied to the beta-series from each run in order to align them across runs. The resulting realigned beta-series then served as input patterns for classification and PSA analysis (described below).

### *Regions of interest*

All analyses were performed in anatomically defined regions of interest (ROIs) that were obtained at the individual participant level using Freesurfer (https://surfer.nmr.mgh.harvard.edu/). These ROIs were primarily chosen to encompass representative visual processing regions and memory-related regions. We included the lateral occipital cortex and the fusiform cortex (the posterior half of Freesurfer-derived fusiform) as visual processing regions. The memory-related ROIS were the hippocampus and the parahippocampal cortex, as well as the inferior and superior parietal cortices that have previously been shown to represent item and category information relevant to memory behaviors (Kuhl and Chun, 2014; Lee and Kuhl, 2016; Lee et al., 2016). The localization of the ROIs in a representative subject are depicted in Fig. 3. Results for individual ROIs are depicted but the differences between trial timing conditions were of primary interest rather than differences between ROIs.

### *Category representation using machine-learning classifiers*

The first goal of the study was to test how the trade-off between the SOA and the number of trials influences the decoding of category-level information (animals/tools) using a machine-learning classifier. We chose these categories because their distinction was expected to be subtle, providing a greater challenge for decoding compared to categories popular for their universal robustness, such as faces vs. scenes. All classification analyses were conducted using a support vector machines classifier (SVM) with a linear kernel as implemented in the PyMVPA toolbox (www.pymvpa.org) for Python (www.python.org). As each design was scanned across two runs, each of the two cross-validation folds used a single run of data to train the classifier before using the other run for testing (Fig. 2**D**). Before training the classifier, we performed a sensitivity-based feature selection. Within training data only, a voxel-by-voxel one-way ANOVA was performed to identify the voxels showing a differential response to the two conditions. The 100 voxels with the strongest differential response to tools and animals (as indexed by the largest F-values) were then used as features for training the classifier. This approach is referred to as no-peaking ANOVA feature selection, as it is blind to whether those same voxels are differentially responsive in the testing set, avoiding bias in the cross-validation. The trained classifier was then tested on the second run of the given design. The roles of the two runs were then switched and the cross-validated performance was averaged across the two cross-validation folds. The resulting classification performance was then averaged across subjects for each trial timing condition and ROI. The effects of the timing condition and ROI were formally assessed using repeated measures ANOVA with timing and region as within-subject factors. Because the effect of ROI was not of particular interest in this study, only the effect of timing is reported in the text.

### *Item representation using PSA*

PSA was used to index the degree of neural item representation as estimated in each trial timing condition. First, we calculated a pattern similarity score for each pair of beta-series patterns across the two runs of the same design (e.g., run1-trial1 with run2-trial1, run1-trial1 with run2-trial2, etc). To obtain the similarity scores within each ROI, we extracted the vectors of activation values for all voxels within that ROI and correlated them using the Spearman's correlation coefficient *r*. As single-trial pattern estimates tend to be noisy due to occasional outlier values arising from motion or partial volume effects, Spearman's rank correlation provides a computationally simple way to reduce possible effects of outlying values on the similarity scores. However, the pattern of results reported here does not depend on the similarity metric used. Pattern similarity was never assessed within a run to avoid confounding the data with local autocorrelations and to focus on the reliability of the estimates across runs. This resulted in 12*12=144 r-values for each design in the item model and a larger, variable number of r-values for each design in the trial model (24*24=576 r-values in the slow designs, 36*36=1296 r-values in the medium design, and 48*48=2304 r-values in the quick designs). The resulting r-values were Fisher z-transformed and sorted into three groups depending on whether they indexed the similarity of patterns evoked by: (1) the same item (e.g., cat from run 1 to cat from run 2), (2) two distinct items from the same category (e.g., cat from run 1 to dog from run 2), or (3) two items from different categories (e.g., cat from run 1 to hammer from run 2). Item

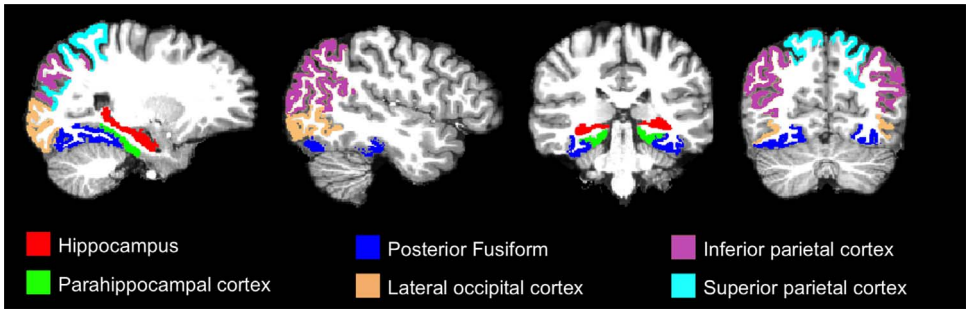


**Fig. 3. Regions of interest from a representative subject.** Regions were identified in the native space of each subject using an automated segmentation in Freesurfer. The average number of voxels across participants in each region was 1150 (sd=101) in the hippocampus, 590 (sd=63) in the parahippocampal cortex, 2128 (sd=331) in the posterior fusiform, 2935 (sd=353) in the lateral occipital cortex, 3807 (sd=491) in the inferior parietal cortex, and 3443 (sd=474) in the superior parietal cortex.

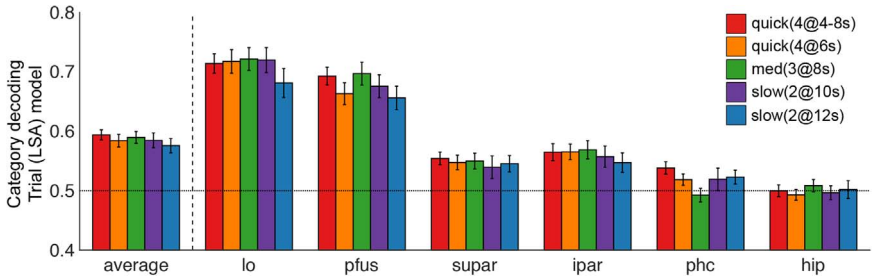


**Fig. 4.** Category-level decoding accuracy from SVM classification. The error bars represent across-subject standard error of the mean. The dashed line represents a theoretical chance of 0.5 (two categories). hip=hippocampus, phc=parahippocampal cortex, pfus=posterior fusiform, lo=lateral occipital cortex, ipar=inferior parietal cortex, supar=superior parietal cortex, average=decoding accuracy averaged across ROIs.

representation indexes were computed for each participant as the normalized mean difference (Cohen's D) between the scores of patterns evoked by the same item vs. patterns evoked by items in the same category across runs (e.g., how much more similar is the cat from run 1 to the same cat from run 2 than to other animals from run 2; Fig. 2**E**). The resulting item representation estimates from the six ROIs were entered into a repeated-measures ANOVA with trial timing and ROI as factors (only the effects of trial timing are reported).

*Item representations and subsequent memory effects*

When an item is repeated, memory for its prior encounters may be reactivated, leading to a reinstatement of the cortical activation pattern present during prior encoding events (Howard and Kahana, 1999; Manning et al., 2011; Polyn et al., 2005). This memory reactivation may contribute to the neural pattern similarity across repetitions of the same item, enhancing item representation. As a consequence, items that are remembered should show more similar neural patterns across repetitions than items that are forgotten. Indeed, several studies have found such memory-based effects on neural pattern similarity (Kuhl and Chun, 2014; Xue et al., 2010). We thus compared memory effects on item representations across trial timing conditions as a third dependent measure in our study. To obtain a memory effect index in each condition, we first sorted the pattern similarity scores for the same items across runs (e.g., cat from run 1 to cat from run 2; dog from run 1 to dog from run 2) based on whether they were subsequently remembered or forgotten, separately for each subject and trial timing condition. Within each condition, we then computed the normalized mean difference (Cohen D) between the Fisher z-transformed similarity scores for remembered vs. forgotten items. Positive scores signal greater pattern similarity for items subsequently remembered than for those forgotten. Subjects that remembered every item in a given condition were excluded from the analysis of that condition. The number of subjects that had at least 1 forgotten trial was 29 in the quick (4@4–8 s), med (3@8 s) and slow (2@10 s) conditions and 25 in the quick (4@6 s) and slow (2@12 s) conditions. The overall memory effect and the interaction between memory and trial timing condition were formally assessed using a repeated-measures ANOVA. For this analysis, participants were excluded list-wise if they were missing forgotten trials in any condition, resulting in 18 subjects included in the ANOVA.

## Results

*Memory performance*

The average proportion of remembered items (Fig. 1**D**) was approximately 80% and did not differ between trial timing conditions (no significant main effect of condition in the repeated-measures ANOVA, $F(4,136)=1.05$, $p=0.381$). There was also no difference in hit rates or false alarm rates between conditions (both $F < 1$, Suppl. Fig. 3). The reaction times (Fig. 1**E**) were marginally different across conditions ($F(4,136)=2.24$, $p=0.068$). Post-hoc pairwise differences showed slower reaction times in the slow (2@10 s) condition than in the med (3@8 s) condition ($t(34)=2.85$, $p=0.007$).

*Category representations via SVM classification*

We first tested the degree to which visual categories (animals vs. tools) can be differentiated from activation patterns in the selected ROIs in each trial-timing condition. The trial (LSA) results, presented in Fig. 4, show that classification was above chance in every condition for several ROIs and on average across ROIs. Classification accuracy did not differ significantly among trial timing conditions (main effect of timing $F < 1$, linear contrast effect $F(1,34)=1.43$, $p=0.24$). The iterative trial (LSS) model provided equivalent results to the trial (LSA) model (Suppl. Fig. 4**,** top panel), including the same overall classification accuracy (LSA grand mean=0.586, se=0.006; LSS grand mean=0.584, se=0.006). The item model yielded an overall greater classification accuracy (grand mean=0.616, se=0.009), with increasing decoding accuracy in the fixed trial-timing designs from slow to quick designs (Suppl. Fig. 4**,** bottom panel).

*Item representations via PSA*

The item-level information estimates from the trial (LSA) model are depicted in Fig. 5. A repeated-measures ANOVA showed an intercept significantly above zero (grand mean=0.065, se=0.006, $F(1,34)=126$, $p < 0.001$), indicating a reliable item representation on average across conditions and ROIs. The magnitude of item representation estimates increased linearly with longer trial timings (main effect of timing: $F(4,136)=2.22$, $p=0.07$, linear increase $F(1,34)=7.28$, $p=0.011$), although it was above chance in all trial timing conditions (lowest mean item representation across ROIs=0.044, se=0.008, $t(34)=5.39$, $p < 0.001$). Post-hoc pair-wise comparisons showed a greater item representation in the slow (2@12 s) design than in the two quick designs (both $t(34) > 2.24$, $p < 0.032$). A similar pattern of results was observed using the trial (LSS) model, with overall item representation estimates from the LSS method marginally greater than the LSA method (Suppl. Fig. 5**,** top panel). The item model yielded significantly greater item representation estimates than the trial (LSA) model, which did not differ between conditions (Suppl. Fig. 5, bottom panel).

*Item representations and subsequent memory effects*

Consistent with a prior report (Xue et al., 2010), we found a reliable effect of memory on item representations (grand mean=0.053, se=0.023, $F(1,17)=5.05$, $p=0.038$; Fig. 6). Marginal differences between conditions ($F(4,68)=2.35$, $p=0.063$) were driven by an advantage for both the slow (2@12 s) and the jittered quick (4@4–8 s) designs (quadratic effect of timing $F(1,17)=4.84$, $p=0.042$), which were the only two designs in which the memory effects were significant on average across ROIs (quick (4@4–8 s) $t(28)=1.8$, $p=0.038$; slow (2@12 s) $t(24)=2.69$, $p=0.006$). Post-hoc pairwise comparisons showed the slow (2@

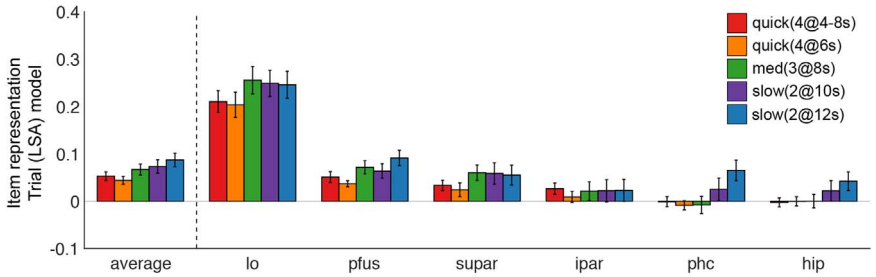


**Fig. 5.** Item-specific pattern similarity across runs. Bar height represents the item representation, indexed as a normalized mean difference (Cohen's D) between the mean neural similarity of the same items across runs to similarity of different items from the same category across runs, averaged across participants. Error bars represent across-subject standard error of the mean.

12 s) design outperformed all shorter fixed trial timing designs (all fixed duration SOA conditions $p < 0.05$, comparison to jittered quick (4@4–8 s) p=0.051). The quick (4@4–8 s) design did not differ significantly from other conditions. The memory effects estimated using the trial (LSS) model did not reach significance or differ between conditions (Suppl. Fig. 6, top panel). The item model showed results comparable to those of the trial (LSA) model with an advantage for the slow trials (Suppl. Fig. 6, bottom panel).

## Discussion

The current gold standards of fMRI design for univariate analyses – slow event-related design and rapid jittered design – are inconvenient from an experimental perspective and may not be optimal for pattern-information analyses. This study empirically tested the effects of different fMRI trial timing designs on the estimability of category-level and item-level neural representations during the encoding of animal and tool images. We compared trial timing conditions with a fixed SOA ranging from a quick design with a 6 s SOA to slow designs with a 12 s SOA, as well as a quick design in which the SOA was jittered. The trial timing conditions further varied in the number of repetitions of each image, so that the scan time was about equal across designs. Therefore, the slow designs fit fewer trials into a run than did quick designs. Results showed that category-level information (tools vs. animal) on a trial-by-trial level was equally detectable using SVM classification in all trial-timing conditions, allowing flexibility in selecting an fMRI trial timing design when category-level information is of primary interest. The detectability of item-level information (cat vs. dog vs. lion) on a trial-by-trial level using PSA increased linearly with increased SOAs, with the slow 12 s SOA design outperforming the quick designs. Finally, the trial-by-trial item similarity across repetitions was greater for items subsequently remembered than for those forgotten, but this effect was best detected with slow 12 s trials. The memory effects were also detectable in the jittered quick design, although they were less pronounced.

We also compared the trial timing conditions using GLM models in which the trial-by-trial information is not of interest and instead representations of individual items are modeled across repetitions. In such models, both category-level and item-level information are better estimated than in trial-by-trial models across all trial timing conditions. The category-level decoding increased linearly from slow to quick for fixed SOA designs while item-level information was equally estimable under all trial timing conditions. These findings demonstrate that designs optimized for the detection of univariate contrasts are not necessarily optimized for multivariate pattern information analyses. Optimal designs may depend on the specific pattern-information analysis of interest and the necessity of decoding the information on a trial-by-trial basis. Furthermore, for many analyses, researchers may have a much wider range of trial timing options than is typically considered.

### *Trial timing and category decoding*

Since the seminal pattern-information decoding paper by Haxby et al. (2001), many studies have purposefully incorporated distinct classes of visual images, such as faces, objects, or scenes, into their fMRI research designs. These studies have shown the robustness of MVPA in decoding the category of the currently viewed content (Haxby et al., 2001; 2011; Kriegeskorte et al., 2008b; O'Toole et al., 2005; Proklova et al., 2016) or even of the retrieved memory content (Gershman et al., 2013; Kuhl et al., 2011; Lewis-Peacock and Norman, 2014; Polyn, 2005) from distributed activation patterns in the visual cortex. The current study further highlights the utility and robustness of the classifier-based decoding approaches, even for subtler category distinctions within the object category. The patterns evoked by the encoding of animal images were reliably differentiable from the patterns evoked by the encoding of tool images on a trial-by-trial basis under all timing conditions in many regions. Within each condition, this was achieved even with just two runs available for cross-validation, corresponding to less than 5 min worth of training data per cross-validation fold. The iterative modeling of individual trials, also known as the LSS method, yielded the same performance as a trial model that included all individual trial regressors in a single model (the LSA method). For both the LSA and the LSS models, we observed no reliable differences between different trial timing conditions on trial-by-trial animal/tool category decoding accuracy. This suggests that

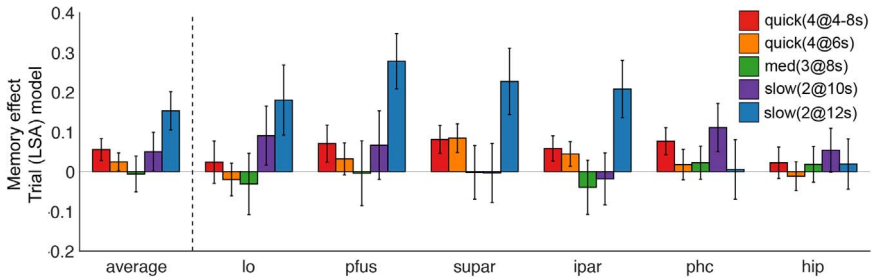


**Fig. 6.** Pattern similarity reflecting subsequent memory, shown by the normalized difference between pattern similarity scores across runs for items subsequently remembered vs. forgotten in each condition. Subjects were excluded on a condition-by-condition basis when they lacked forgotten trials in that condition. Error bars denote across-subject standard error of the mean.

researchers planning to employ classifier-based approaches to visual category decoding may have a range of options available and can base their choice on other constraints, such as the psychological suitability of the design or a plan to also estimate univariate effects. When trial-by-trial information is not of interest, modeling all the repetitions of the same item in a single regressor improves category decoding compared to trial-by-trial models. In such cases, the quick trial timing with a fixed 6 s SOA provided better category decoding than the slow designs. Therefore, in such studies, long SOAs may be sacrificed in favor of increased number of trials.

*Trial timing and item-level representations*

The current study also demonstrated the detectability of item-level information, indexed as the normalized difference between across-runs pattern similarity for the same item and the pattern similarity of different items from the same category. The item-level information on a trial-by-trial level was best detected with fewer, slower trials irrespective of whether individual trials were modeled via the LSA or LSS method. Overall, the iterative LSS method provided a marginally greater estimate of item representations. When item-level information at the trial level was not of interest, using the item model provided a significant boost in item-level estimates, with no significant differences among conditions.

The trial-by-trial pattern similarity scores of individual items across their repetitions were greater for subsequently remembered items than for those forgotten. The memory effects were best detected using slow trials with 12 s onset-to-onset timing. These results are consistent with prior work showing that slow trials are necessary to detect trial-by-trial differences in item representations across learning (Visser et al., 2016). Therefore, experiments planning to use PSA on single trial level or experiments constrained to a single presentation per item may benefit from slow trial designs, even if it means fewer trials overall. Importantly, however, the trial-by-trial item-level information was detectable in the present study in most regions in each trial-timing condition. Furthermore, the memory effects were observed not only with slow trials but also in the jittered quick design. Thus, faster trials do not preclude item-specific decoding.

Interestingly, the memory effects on item representations were numerically strongest in the LSA model. These findings are in line with prior modeling work that showed an advantage of the LSA model when the trial-by-trial variability of the neural response is high or is of interest (Abdulrahman and Henson, 2016). While models that include multiple events in a single regressor typically provide an advantage in terms of more stable and smoother parameter estimates (Abdulrahman and Henson, 2016), they ignore the potential trial-by-trial variability, which ends up in the error term. In this case, items that are subsequently forgotten may not be recognized as being the same across repetitions, and may be represented somewhat differently on each trial, lessening the sensitivity of item models that aggregate across repetitions. For similar reasons, LSS models may not capture all the variance arising from closely proximal trials in the combined regressor, and either erroneously attribute that variance to the trial that is currently being modeled by its own regressor, or leave that variance unexplained in the error term. Modeling each trial via LSA may provide the most sensitive method for capturing trial-by-trial effects (Abdulrahman and Henson, 2016). Alternatively, a less restrictive version of the LSS model, such as one that splits the remaining trials based on conditions rather than combining them in a single regressor, may also avoid some of the problems associated with uncaptured variability (Turner et al., 2012).

*Different GLM approaches and pattern-information analyses*

Our first interest was to determine the degree to which category information and item-level information are detectable on a single-trial basis. We have taken two approaches to estimating the trial-by-trial activation patterns: LSA, which models all trials by a separate regressor within a single GLM model, and LSS, which iteratively models one trial by a separate regressor and all other trials by a second regressor to improve the stability of the single-trial estimates (Mumford et al., 2012). As noted above, we found no difference between these methods for category decoding, a marginal LSS advantage for detecting item-level information, and a numerical LSA advantage for detecting memory effects that exploits the variability in trial-by-trial estimates across repetitions. The relative benefit of LSA vs. LSS for different analyses differed minimally across the trial timing conditions. Thus, within the timing range considered here, the choice of GLM model for obtaining single trial estimates (LSA vs. LSS) should be primarily driven by the expected trial-by-trial variability and the effects that the pattern-information analysis aims to capture (Abdulrahman and Henson, 2016).

Trial timing seems important, however, for decisions regarding whether or not to include temporal derivatives into GLM models. Here, including temporal derivatives in the quick conditions, with SOAs as low as 4 or 6 s, yielded worse univariate and multivariate results than when temporal derivatives were left out. In contrast, including temporal derivatives in conditions that had SOAs of at least 8 s improved the same results (compared to the GLM versions without temporal derivatives), even in the trial-by-trial LSA model. Adding temporal derivatives into GLM models has been shown to generally improve model fits by accounting for small hemodynamic response lag differences (Calhoun et al., 2004; Friston et al., 1998; Worsley and Taylor, 2006), but may add collinearity to condition-rich designs with closely spaced trials. Our results provide new insights into how this trade-off affects GLM estimates under varying trial timing scenarios that can be used to inform future investigations. Please note that following prior research (Mumford et al., 2012, 2014), we did not include temporal derivatives when obtaining single trial estimates using the LSS method, but it is in principle possible and could be beneficial with slower trials.

Our data also suggest that the advantage for detecting item-level information using a slow design is not fully explainable by differences in the number of regressors that need to be estimated with a limited number of observations (time points) in a run. First, the trial (LSS) models and the item models have the same number of regressors in each trial timing condition (the LSS model had only two task-related regressors+motion parameters in each iteration, the item model always had 12 task regressors+motion parameters). Yet, the trial LSS model showed a similar slow design advantage as the LSA model in item representation estimates and the item model showed a similar slow design advantage in detecting memory effects. Additionally, the trial LSA model does scale with the number of trials. However, the number of regressors ended up equal between quick conditions (without temporal derivatives) and slow conditions (with temporal derivatives). Thus, the onset-to-onset timing of trials may be a stronger boundary condition for detecting item-level information than the number of regressors in GLM models per se.

*The effect of jitter on pattern-information analyses*

For standard univariate analyses with few conditions, jittered trial onsets enable a faster trial timing and a better contrast detection than regularly spaced trials (Burock et al., 1998). One question of the current study was whether SOA jittering is equally useful for pattern-information analyses and condition-rich designs, or whether a regular onset-to-onset trial timing may yield similar results. We reasoned that the benefits of jitter are limited in condition-rich designs, for example because the random ordering of trials from different conditions may already ensure an uncorrelated predicted BOLD response for each condition. Furthermore, jitter may introduce some psychological variability between long and short trials that would be reflected in

the BOLD activation patterns, adding noise to the category and item representation estimates. As prior studies have shown that single-trial estimates are unlikely to be reliable with trial timing less than 4–6 s (Abdulrahman and Henson, 2016; Mumford et al., 2012; Visser et al., 2016), our study compared two "quick" conditions with an equal number of trials that were either jittered with an SOA of 4 or 8 s (Kriegeskorte et al., 2008b) or spaced regularly every 6 s. We found no significant pair-wise differences between these conditions in any analyses. For the trial-by-trial estimates, the jittered version of the quick design showed a numerical advantage compared to the non-jittered version across analyses. For the item-by-item estimates with a single regressor across item repetitions, the non-jittered version provided a numerical advantage. Thus, our data suggest that jitter may not be as crucial in condition-rich designs as it is for standard univariate analyses and researchers may consider whether or not jitter is likely to provide a benefit in a given study. Factors such as possible behavioral effects of jitter or plans to also conduct univariate analyses may tip the scales one way or the other.

### *Limitations*

Pattern-information analyses have been applied to a wide range of questions and the scope of the current study is limited in comparison. For one, to investigate the reliability of category- and item-level pattern-information analyses in different fMRI designs, this study used images of common items from well-learned categories. A wide range of studies implementing pattern-information analyses have relied on principally similar categories and stimuli and thus may be informed by the current study when planning their design. However, pattern-information analyses have also been applied to tracking the changes in representation of new artificial stimuli with arbitrary relations (Schlichting et al., 2015), decoding non-sensory information such as intentions (Haynes et al., 2007) or decoding process-level information such as whether a person is engaging in encoding or retrieval on a given trial (Richter et al., 2016). It is unclear whether findings from the current study are fully generalizable to such applications. Furthermore, the advantage of the slow trials for trial-by-trial item-level decoding in the present study is likely due to both fMRI factors such as overlapping hemodynamic responses to closely spaced trials that cannot be fully differentiated on a trial-by-trial basis as well as cognitive factors, such as a possibly shallower encoding during short trials. Thus, the relative advantage of slow trials may diminish in other cognitive paradigms where the SOA has a smaller impact on cognitive processing.

Prior research (Stark and Squire, 2001) showed that the direction and amplitude of a univariate signal may be affected by an inclusion of a baseline task—a filler task aimed to control cognitive processes, such as mind-wandering, in between trials of interest. We used a passive baseline (passive viewing of a fixation cross) during the inter-stimulus intervals. While a manipulation of the baseline task was beyond the scope of the current study, it would be interesting to test the degree to which an inclusion of a baseline task may affect pattern-information analyses. It is, however, worth noting that an unstructured null time is unlikely to be strongly detrimental to the decoding approaches tested in the current study. If it were, its effect would presumably be worst in the slow design with long inter-stimulus intervals. However, this design proved to be the most reliable for estimating item-level information on a trial-by-trial basis.

One of our dependent measures—the memory effects on pattern similarity across repetitions—was less powered than our other two measures as several participants remembered all the items in some of the conditions. Despite the decreased power, the memory effects were reliable overall. The observed memory effects, however, suggest that it may be impossible to assess the reliability of item representation estimates that would not be influenced by memory. The repetition of an item may trigger reinstatement of a prior encounter with the item and its associated context (Howard and Kahana, 1999; Manning et al., 2011; Polyn et al., 2005). This memory-based reinstatement would contribute to the neural pattern similarity for the same items across repetitions, as evidenced by the significant memory effects obtained in the current study. Thus, item representation estimates across the brain are likely influenced by item memory effects, even in studies that do not explicitly focus on memory.

Despite these limitations, we believe that our findings are informative beyond the fields of visual category representation or memory. They suggest the possibility of successfully employing a wider range of trial timings when designing an fMRI study than is typically considered and highlight the factors that should influence trial timing choices. Our results also provide converging evidence for the advantage of slow trials to detect trial-by-trial variability across events that reflects learning (Visser et al., 2016). The comparison of different GLM approaches (LSS vs. LSA, inclusion vs. exclusion of temporal derivatives) may also inform future GLM choices irrespective of the research question.

### *Conclusions*

The current study investigated the optimal balance between the number of trials and the SOA within a fixed duration scan for pattern-information analyses. For many analyses, category-level and item-level information were similarly detectable across all trial timing conditions. When differences were detected, category decoding benefited from a larger number of regularly spaced trials while slow 12-s trials allowed for a better detection of item-level information and memory status of items, especially on a single trial level. The jittered and non-jittered versions of the quick trial timing performed similarly on all metrics. Notably, category-level and item-level information were, in the current study, detectable to some degree in most conditions, with less than ten minutes of scan time per condition, further highlighting the utility of pattern-information analyses in cognitive neuroscience research. The results may inform the choice of trial timing and analytical approaches when considering decoding analyses in condition-rich designs.

## Acknowledgements

This work was made possible through generous support from the Lewis Family Endowment to the UO which supports the Robert and Beverly Lewis Center for Neuroimaging. We thank Lydia K. McNiel for help with data collection and Caitlin R. Bowman for helpful comments on the manuscript.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.neuroimage.2017.04.025.

## 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. http://dx.doi.org/10.1016/j.neuroimage.2015.11.009.

Antrobus, J.S., Singer, J.L., Greenberg, S., 1966. Studies in the stream of consciousness: experimental enhancement and suppression of spontaneous cognitive processes. Percept. Mot. Skills 23, 399–417.

Buckner, R.L., Bandettini, P.A., O'Craven, K.M., Savoy, R.L., Petersen, S.E., Raichle, M.E., Rosen, B.R., 1996. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. Proc. Natl. Acad. Sci. 93 (25), 14878–14883.

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. http://dx.doi.org/10.1016/j.neuroimage.2003.12.029.

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.

http://dx.doi.org/10.1016/j.neuroimage.2012.03.076.
Dale, A.M., 1999. Optimal experimental design for event-related fMRI. Hum. Brain Mapp. 8 (2–3), 109–114.
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. http://dx.doi.org/10.1006/nimg.1997.0306.
Friston, K.J., Zarahn, E., Josephs, O., Henson, R.N., Dale, A.M., 1999. Stochastic designs in event-related fMRI. NeuroImage 10 (5), 607–619. http://dx.doi.org/10.1006/nimg.1999.0498.
Gershman, S.J., Schapiro, A.C., Hupbach, A., Norman, K.A., 2013. Neural context reinstatement predicts memory misattribution. J. Neurosci.: Off. J. Soc. Neurosci. 33 (20), 8590–8595. http://dx.doi.org/10.1523/JNEUROSCI.0096-13.2013.
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. http://dx.doi.org/10.1126/science.1063736.
Haxby, J.V., Guntupalli, J.S., Connolly, A.C., Halchenko, Y.O., Conroy, B.R., Gobbini, M.I., et al., 2011. A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron 72 (2), 404–416. http://dx.doi.org/10.1016/j.neuron.2011.08.026.
Haynes, J.-D., Sakai, K., Rees, G., Gilbert, S., Frith, C., Passingham, R.E., 2007. Reading hidden intentions in the human brain. Curr. Biol. 17 (4), 323–328. http://dx.doi.org/10.1016/j.cub.2006.11.072.
Henson, R., 2007. CHAPTER 15 efficient experimental design for fMRI. In: Statistical Parametric Mapping the Analysis of Functional Brain Images. Elsevier, pp. 193–210. http://doi.org/10.1016/B978-012372560-8/50015-2.
Howard, M.W., Kahana, M.J., 1999. Contextual variability and serial position effects in free recall. J. Exp. Psychol.: Learn. Mem. Cogn. 25 (4), 923–941.
Jimura, K., Poldrack, R.A., 2012. Analyses of regional-average activation and multivoxel pattern information tell complementary stories. Neuropsychologia 50 (4), 544–552. http://dx.doi.org/10.1016/j.neuropsychologia.2011.11.007.
Kriegeskorte, N., Bandettini, P., 2007. Analyzing for information, not activation, to exploit high-resolution fMRI. NeuroImage 38 (4), 649–662. http://dx.doi.org/10.1016/j.neuroimage.2007.02.022.
Kriegeskorte, N., Mur, M., Bandettini, P., 2008a. Representational similarity analysis – connecting the branches of systems neuroscience. Front. Syst. Neurosci., 1–28. http://dx.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. http://dx.doi.org/10.1016/j.neuron.2008.10.043.
Kuhl, B.A., Chun, M.M., 2014. Successful remembering elicits event-specific activity patterns in lateral parietal cortex. J. Neurosci.: Off. J. Soc. Neurosci. 34 (23), 8051–8060. http://dx.doi.org/10.1523/JNEUROSCI.4328-13.2014.
Kuhl, B.A., Rissman, J., Chun, M.M., Wagner, A.D., 2011. Fidelity of neural reactivation reveals competition between memories. Proc. Natl. Acad. Sci. USA 108 (14), 5903–5908. http://dx.doi.org/10.1073/pnas.1016939108.
Lee, H., Kuhl, B.A., 2016. Reconstructing perceived and retrieved faces from activity patterns in lateral parietal cortex. J. Neurosci.: Off. J. Soc. Neurosci. 36 (22), 6069–6082. http://dx.doi.org/10.1523/JNEUROSCI.4286-15.2016.
Lee, H., Chun, M.M., Kuhl, B.A., 2016. Lower parietal encoding activation is associated with sharper information and better memory. Cereb. Cortex, bhw097. http://dx.doi.org/10.1093/cercor/bhw097.
Lewis-Peacock, J.A., Norman, K.A., 2014. Competition between items in working memory leads to forgetting. Nat. Commun. 5, 5768. http://dx.doi.org/10.1038/ncomms6768.
Manning, J.R., Polyn, S.M., Baltuch, G.H., Litt, B., Kahana, M.J., 2011. Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. Proc. Natl. Acad. Sci. USA 108 (31), 12893–12897. http://dx.doi.org/10.1073/pnas.1015174108.
McCarthy, G., Luby, M., Gore, J., Goldman-Rakic, P., 1997. Infrequent events transiently activate human prefrontal and parietal cortex as measured by functional MRI. J. Neurophysiol. 77 (3), 1630–1634.
Mumford, J.A., Davis, T., Poldrack, R.A., 2014. The impact of study design on pattern estimation for single-trial multivariate pattern analysis. NeuroImage 103, 130–138. http://dx.doi.org/10.1016/j.neuroimage.2014.09.026.
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. http://dx.doi.org/10.1016/j.neuroimage.2011.08.076.
Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci. 10 (9), 424–430. http://dx.doi.org/10.1016/j.tics.2006.07.005.
O'Toole, A.J., Jiang, F., Abdi, H., Haxby, J.V., 2005. Partially distributed representations of objects and faces in ventral temporal cortex. J. Cogn. Neurosci. 17 (4), 580–590. http://dx.doi.org/10.1162/0898929053467550.
Pereira, F., Mitchell, T., Botvinick, M., 2009. Machine learning classifiers and fMRI: a tutorial overview. NeuroImage 45 (S1), S199–S209. http://dx.doi.org/10.1016/j.neuroimage.2008.11.007.
Polyn, S.M., 2005. Category-specific cortical activity precedes retrieval during memory search. Science 310 (5756), 1963–1966. http://dx.doi.org/10.1126/science.1117645.
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. http://dx.doi.org/10.1126/science.1117645.
Proklova, D., Kaiser, D., Peelen, M.V., 2016. Disentangling representations of object shape and object category in human visual cortex: the animate-inanimate distinction. J. Cogn. Neurosci. 28 (5), 680–692. http://dx.doi.org/10.1162/jocn_a_00924.
Richter, F.R., Chanales, A.J.H., Kuhl, B.A., 2016. Predicting the integration of overlapping memories by decoding mnemonic processing states during learning. NeuroImage 124 (Pt A), 323–335. http://dx.doi.org/10.1016/j.neuroimage.2015.08.051.
Schlichting, M.L., Mumford, J.A., Preston, A.R., 2015. Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. Nat. Commun. 6, 1–10. http://dx.doi.org/10.1038/ncomms9151.
Stark, C.E., Squire, L.R., 2001. When zero is not zero: the problem of ambiguous baseline conditions in fMRI. Proc. Natl. Acad. Sci. 98 (22), 12760–12766. http://dx.doi.org/10.1073/pnas.221462998.
Turner, B.O., Mumford, J.A., Poldrack, R.A., Ashby, F.G., 2012. Spatiotemporal activity estimation for multivoxel pattern analysis with rapid event-related designs. NeuroImage 62 (3), 1429–1438. http://dx.doi.org/10.1016/j.neuroimage.2012.05.057.
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. http://dx.doi.org/10.1111/psyp.12665.
Ward, E.J., Chun, M.M., Kuhl, B.A., 2013. Repetition suppression and multi-voxel pattern similarity differentially track implicit and explicit visual memory. J. Neurosci.: Off. J. Soc. Neurosci. 33 (37), 14749–14757. http://dx.doi.org/10.1523/JNEUROSCI.4889-12.2013.
Worsley, K.J., Taylor, J.E., 2006. Detecting fMRI activation allowing for unknown latency of the hemodynamic response. NeuroImage 29 (2), 649–654. http://dx.doi.org/10.1016/j.neuroimage.2005.07.032.
Xue, G., Dong, Q., Chen, C., Lu, Z., Mumford, J.A., Poldrack, R.A., 2010. Greater neural pattern similarity across repetitions is associated with better memory. Science 330 (6000), 97–101. http://dx.doi.org/10.1126/science.1193125.