NeuroImage 148 (2017) 197–200

Contents lists available at ScienceDirect

NeuroImage

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

Editorial

Avoiding illusory effects in representational similarity analysis: What (not) to do with the diagonal

ARTICLE INFO

Keywords:
Neuroimaging
Representational similarity analysis
Object categorization
Ventral temporal cortex

ABSTRACT

Representational similarity analysis (RSA) is an important part of the methodological toolkit in neuroimaging research. The focus of the approach is the construction of representational dissimilarity matrices (RDMs), which provide a single format for making comparisons between different neural data types, computational models, and behavior. We highlight two issues for the construction and comparison of RDMs. First, the diagonal values of RDMs, which should reflect within condition reliability of neural patterns, are typically not estimated in RSA. However, without such an estimate, one lacks a measure of the reliability of an RDM as a whole. Thus, when carrying out RSA, one should calculate the diagonal values of RDMs and not take them for granted. Second, although diagonal values of a correlation matrix can be used to estimate the reliability of neural patterns, these values must nonetheless be excluded when comparing RDMs. Via a simple simulation we show that inclusion of these values can generate convincing looking, but entirely illusory, correlations between independent and entirely unrelated data sets. Both of these points are further illustrated by a critical discussion of Coggan et al. (2016), who investigated the extent to which category-selectivity in the ventral temporal cortex can be accounted for by low-level image properties of visual object stimuli. We observe that their results may depend on the improper inclusion of diagonal values in their analysis.

Representational similarity analysis (RSA) is an important part of the multivariate pattern analysis (MVPA) toolkit. The utility of RSA arises from its focus on the construction of representational dissimilarity matrices (RDMs), which provide a common medium for making direct comparisons between various neural data types (cellular recordings, fMRI, and M/EEG), computational models, and behavior (Kriegeskorte et al., 2008a).

A dissimilarity matrix as used in RSA is based on the mathematical notion of a distance matrix, which consists of the pairwise distances (or differences) between each element of a set of values. For example, if we imagine a simple 2D space, a typical distance matrix will consist of the pairwise Euclidean distances between all points in the space (Fig. 1A). The value of a diagonal cell in such a matrix reflects the distance between an element and itself, and by definition, is 0. Thus there is an important difference in type between the on-diagonal and off-diagonal cells in a distance matrix. This type-difference is also crucial for RDMs, in which several measures are available for computing pairwise dissimilarity (Kriegeskorte et al., 2008a; Nili et al., 2014; Walther et al., 2016). It is also essential for related visualization and dimensionality techniques that can be applied to RDMs. For example, having 0 values along the diagonal is required for standard forms of multidimensional scaling. While analysis with RDMs is in principle simple, it is deceptively so. This is especially the case when dissimilarity is computed using $1 - r$, where r is the Pearson correlation between (say) the neural patterns for two stimulus conditions in some ROI as measured with fMRI (see e.g. Kriegeskorte et al., 2008b). Here we highlight two issues pertaining to the diagonal that can arise with the construction and comparison of RDMs using $1 - r$, which is the most commonly used measure for dissimilarity in applications of RSA.

The first issue is that when constructing a neural RDM one cannot take the difference between the on-diagonal and off-diagonal cells for granted. It is not uncommon to compute the dissimilarity, as $1 - r$, for all off-diagonal cells of a neural RDM, assuming that for the on-diagonal cells all neural patterns must be identical to themselves. However correlations also provide a means of testing for the reliability and discriminability of activation patterns for an experimental condition (Haxby et al., 2001). Thus if one wishes to carry out RSA, one should first compute the within (on-diagonal) and between (off-diagonal) similarities using correlations or some other metric. If a pattern is reliable—that is, it contains information about the experimental condition—then the on-diagonal (within-condition) correlation should generally be higher than any of the off-diagonal values in the matrix (Haxby et al., 2001). However, if the on-diagonal correlations are not higher, then there is no reliable difference between the neural patterns for the different conditions, and there is no type-difference between the on- and off-diagonal elements of the matrix. In other words, without reliable on-diagonal similarity, there is no reliable off-diagonal dissimilarity, so it makes little analytic sense to construct an RDM for further analysis.

Crucially on-diagonal correlations can only be computed by partitioning the data into subsets, so that two independent measurements for each condition can be correlated. In many implementations of RSA, this is not done. This is not necessarily a problem, since researchers can find other ways to measure the reliability of their matrices by analyzing the replicability of the off-diagonal elements. However, these alternatives will depend strongly upon the existence of variation in similarity between conditions, which is not guaranteed even with reliable neural patterns (for example, if

http://dx.doi.org/10.1016/j.neuroimage.2016.12.079
Received 15 August 2016Accepted 28 December 2016
Available online 07 January 2017

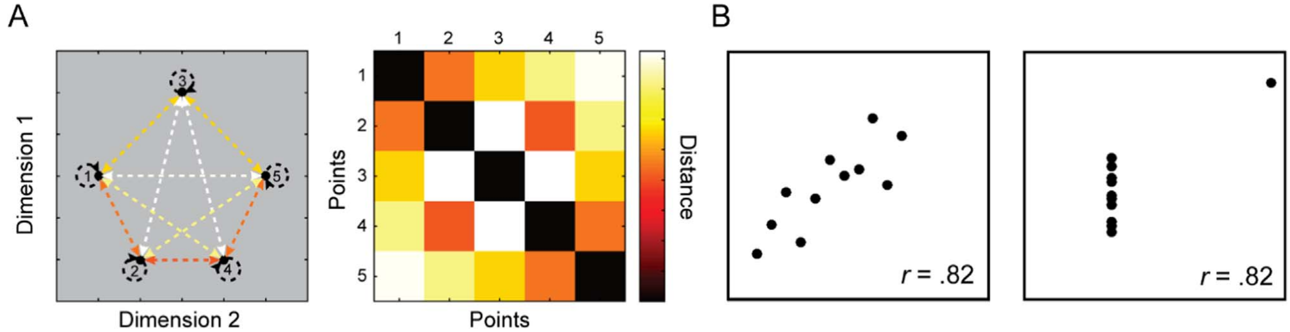

Fig. 1. Distance matrices and correlation scatter plots. (A) A 2D space containing 5 points and the corresponding distance matrix. The color coding of lines (2D space) and cells (matrix) indicate the Euclidean distances between points, with black rings indicating the distance of a point to itself (i.e. 0). (B) Reproduction of Fig. 1 and 4 from Anscombe (1973), depicting two sets of data with radically different scatter plots, but identical correlations (Pearson's r) up to the 2nd decimal point.

the patterns for all conditions are equidistant from each other in activation space). Instead, comparing on- and off-diagonal dissimilarity is a more direct measure of reliability as a first step before analyzing the off-diagonal values (e.g. Bracci and Op de Beeck, 2016; Op de Beeck et al., 2008). Importantly, differences in the on-diagonal values may also contain important information, and can be the subject of analysis (e.g. Harel et al., 2014). Still, for many applications, it is important to estimate the within-condition reliability of neural patterns as a first step, even though the diagonal is excluded in further analysis.

This brings us to the second issue, which is that typically when comparing RDMs *one must exclude the diagonal*. Consider some of the illustrative data of Anscombe (1973). When computing a correlation between the elements of two RDMs, it is easy to imagine something similar to the first plot in Fig. 1B, in which there is a stable, positive relationship between values. However when including the diagonal, the relationship might also be similar to the second plot, in which the correlation is identical to the first plot, but in which the effect is driven by an extreme outlier. The reason is that if the on-diagonal vales are reliably less dissimilar (i.e. the within-condition correlations are higher), then they are effectively outliers relative to the distribution of off-diagonal values. Recall that there is a type difference between the on- and off-diagonal cells of a distance matrix making them not directly comparable. When on-diagonal values are included in a correlation this creates bimodal distributions of correlated values, which can result in robust correlations driven entirely by noise.

To illustrate the above point, we ran a simple simulation (Fig. 2A). Two independent sets of binary "neural patterns" were randomly generated reflecting two completely different experiments (A and B), each containing N different stimulus conditions (s_1,...,s_N), for N=5, 10, 20, or 40. To ensure there was some positive, but weak, relationship between the patterns within an experiment, 60% of bits were flipped to be identical across stimulus conditions. To simulate a test for within- and between-condition pattern reliability, 20–35% of the bits for each pattern were then independently flipped again, to create a second set of patterns for each stimulus condition, as one might generate from halving and averaging neural patterns across runs of an fMRI experiment (cf. Haxby et al. 2001). The above process was carried out independently for the patterns from

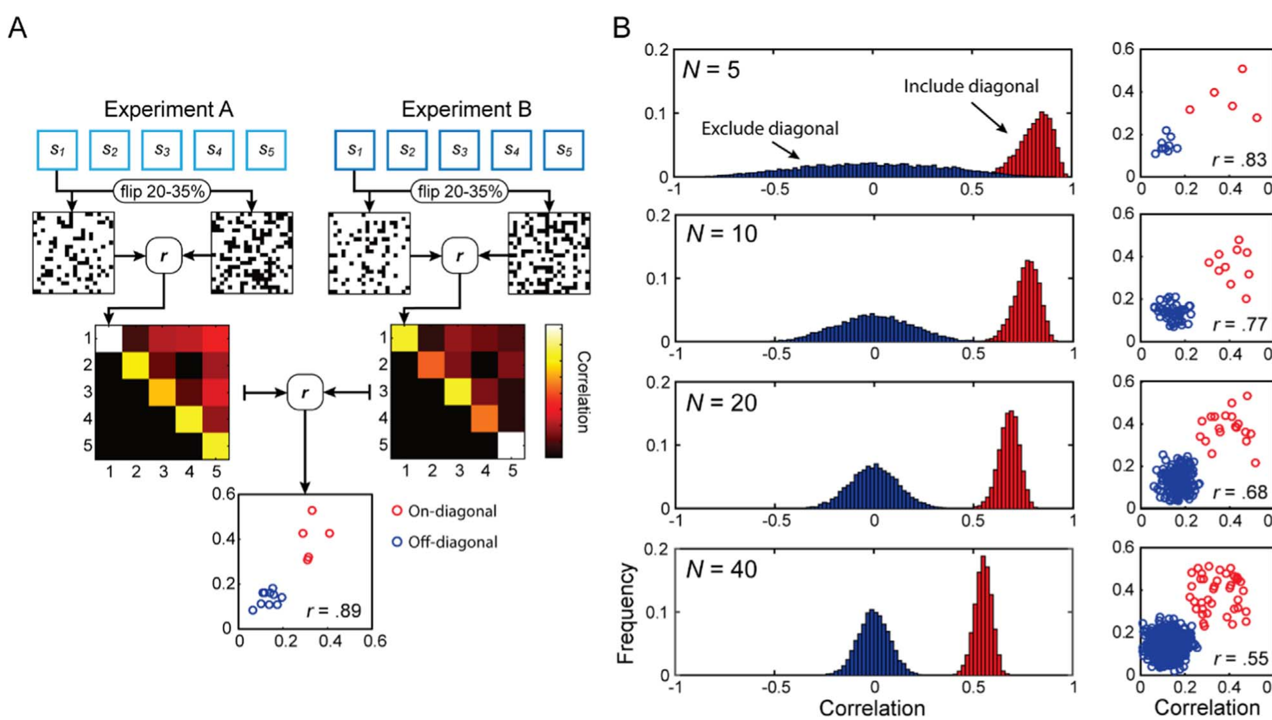

Fig. 2. Simulation methods and results. (A) Depiction of one iteration of the simulation (N=5), including pairs of stimulus "halves" for s_1 in experiments A and B, as well as the complete correlation similarity matrices (CSMs) and resulting correlation and scatter plot (note that the correlation is calculated on all points, blue and red). (B) histograms of correlations both excluding (blue) and including (red) the on-diagonal values, for each value of N (bin size=.02, for all histograms). Scatter plots are for the correlation that was closest to the median of the on-diagonal 'red' correlation distributions. Off- and on-diagonal values in the scatter plots are coded blue and red respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)