Summary of my vision-related research

Rajeev Raizada
Dept. of Brain & Cognitive Sciences, University of Rochester
http://raizadalab.org

1 Overview of research: structure of neural representations

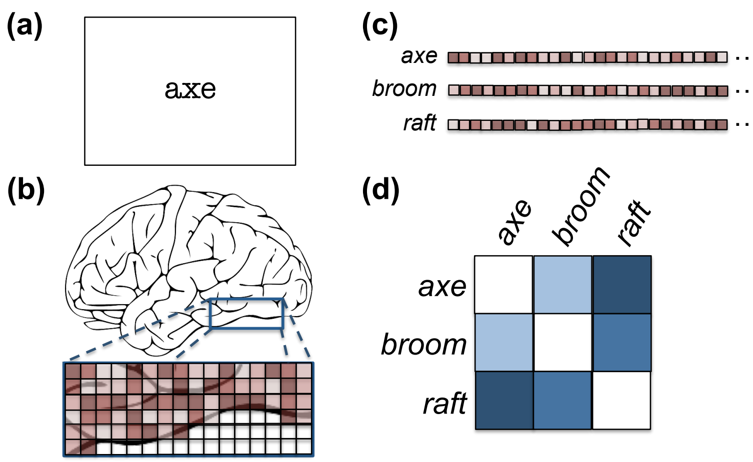

Figure 1: A schematic illustration of how neural similarity can be calculated, and how it can be related to the similarity of items in the world. **(a)** Subjects are presented with a stimulus inside the fMRI scanner, in this case the written word "axe". **(b)** The multivoxel neural activation pattern elicited by that stimulus is measured in some region-of-interest in the brain. **(c)** The activation patterns elicited by different stimuli can be represented as lists of numbers (called vectors), and the similarity between those vectors can be calculated simply as the correlation between them. **(d)** The set of similarities between the neural representations of the different stimuli collectively form a similarity space, here represented as a similarity matrix. The colour of each cell in the matrix represents the neural similarity between the corresponding pair of stimulus words.

Research in my lab combines human neuroimaging, machine learning and behavioural testing in order to explore how people's neural representations are structured. The fundamental driving hypothesis is that the structure of people's internal neural representations should be a function of the external structure of stimuli in the world and of the task being performed. Neural representational structure is measured by computing the pattern-similarity of multivoxel fMRI activation. Some key elements of the overall approach are illustrated in Figure 1. We specifically focus on two aspects of this problem: linguistic representations, and representations of visual form.

The use of machine learning algorithms is central to my work. These algorithms excel at finding structure in distributed patterns of brain activation. Having found that structure, we can then relate those neural activation patterns to people's behaviour.

Such an approach is an instance of the broader term "Data Science", and indeed my lab is closely affiliated with the University of Rochester's new Data Science Institute. I have been appointed as one of the Distinguished Researchers of the New York State Center for Excellence in Data Science (headed by Henry Kautz and Scott Steele), and I recently taught a module in the new cross-departmental Data Science class DSC/BCS 530.

The relation between machine learning and neuroscience in my lab extends beyond just using those algorithms to process fMRI data. We also ask the question of whether algorithms in machine learning and the algorithms that are actually used by the brain may converge onto similar solutions: both are faced with the problem of extracting meaning from noisy and complex input. These potential parallels are most directly explored in my work on visual processing, described below, which investigates kernel-based learning and the representation of manifolds. I also pursue these same issues in research on the structure of the brain's linguistic representations, but in the current document I will describe only my vision-related work.

2 Representations of visual form

In the domain of high-level vision, my lab and I are investigating the representation of visual form. This work is starting to reveal striking parallels between computational strategies that have been developed in machine learning and those that may be actually used by the brain.

2.1 How does the brain learn to solve a linearly inseparable task?

Many real-world categorisation tasks have the property of being linearly inseparable, meaning that it is impossible to separate the two categories using a straight line. Such tasks are computationally challenging, but the brain is able to solve them with apparent ease. In this study, carried out with BCS graduate student Dave Kleinschmidt, former BCS postdoc (now Princeton tenure-track faculty) Lauren Emberson, my lab-manager Donias Doko and Cornell University faculty-member Shimon Edelman, we trained human participants to solve a linearly inseparable 3D shape recognition task, and used fMRI to see investigating their neural representations of the task space were transformed over the course of learning. By measuring the structure of their neural representational spaces, we found that people's brains transformed the task space in a manner remarkably similar to that carried out by algorithms in machine learning known as kernels, in which representing items in terms of their similarities to each other allows linearly inseparable tasks to be transformed into becoming separable. This creates a new bridge between machine learning and human neural information processing, and suggests that both may have converged upon similar solutions to a challenging computational problem.

It has long been recognised that linearly inseparable categorisation problems are more difficult to solve than are ones that are separable, as a neuron which calculates a thresholded weighted sum of its inputs is in effect drawing a linear decision boundary. The classic example of a linearly inseparable problem is the exclusive-or (XOR) task, and this task has formed the basis of much subsequent work. However, the XOR task is far from the only linearly inseparable task, and indeed is arguably not the most natural or ecologically valid one. A very common task structure is one which we may call the "cloud within a ring" problem, in which one category is surrounded on all sides by another. Examples include "members of my own tribe, versus everybody else", or "close to home, versus far from home".

2.2 Kernels, similarity, and linearising a task

Kernels are tools in machine learning which recast the task space into a different form. Although they are often described using somewhat opaque terms such as infinite-dimensional Reproducing Kernel Hilbert Spaces, in essence what kernels do is that they represent a set of data points in terms of their similarities to each other. My colleagues and I recently published a paper in the Journal of Mathematical Psychology, highlighting these under appreciated commonalities between kernels and similarity based representations (Shahbazi et al., 2016).

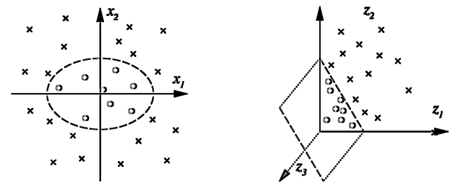

Figure 2: An example of how a kernel can transform a linearly inseparable space into a separable one. This is Figure 2.1 from Schölkopf and Smola (2002).

By transforming a set of data points into the set of similarities between those data points, a problem space which is originally linearly separable can be transformed into being separable. A visual illustration of this is shown in Figure 2, from Schölkopf and Smola (2002). In that instance, the effect of the transformation is that items become represented in terms of their distance from the center. Thus, representing items in terms of their similarities with each other can linearise a problem.

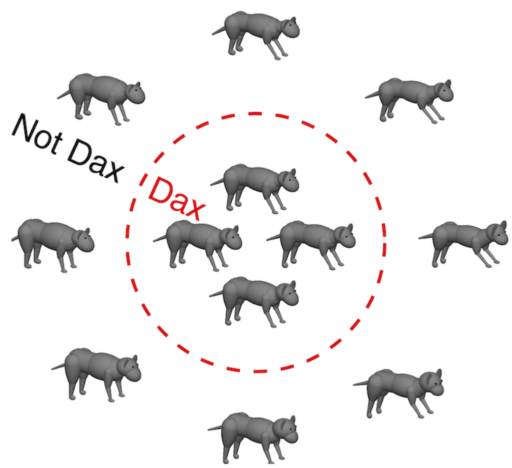

Figure 3: The set of 3D shapes used in our task. The subjects' task was to learn to distinguish animals in the inner ring ("daxes") from those in the outer ring ("non-daxes").

Based on the reasoning above, we created a task with an "cloud within a ring" structure, i.e. in which there were two categories to be distinguished, with one wrapped around the other. The stimuli were morphable 3D animal shapes based on those used in Edelman (1995). The subjects' task was to learn to distinguish animals in the inner ring ("daxes") from those in the outer ring ("non-daxes"), as shown in Figure 3. Over the course of four training sessions, people successfully learned the task. They underwent fMRI scans before and after the training.

2.3 Transformation of neural representational space by learning

One of the key results from this study is illustrated in Fig. 4, in which neural representations corresponding to different rings in the animal shape-space are shown before and after training. Before training, in the left panel, the different parts of shape-space are highly entangled with each other. However, after training, they have become disentangled, such that the parts of shape-space corresponding to the "dax" category are now separable from those corresponding to the non-daxes. This transformation is remarkably parallel to the manner in which a kernel's similarity-based representation linearises an otherwise inseparable task, as shown in Fig. 2.

The paper describing these results (Kleinschmidt et al., 2016) is undergoing final revisions, and will be submitted within the next couple of weeks. We are excited about this result, as it is the first

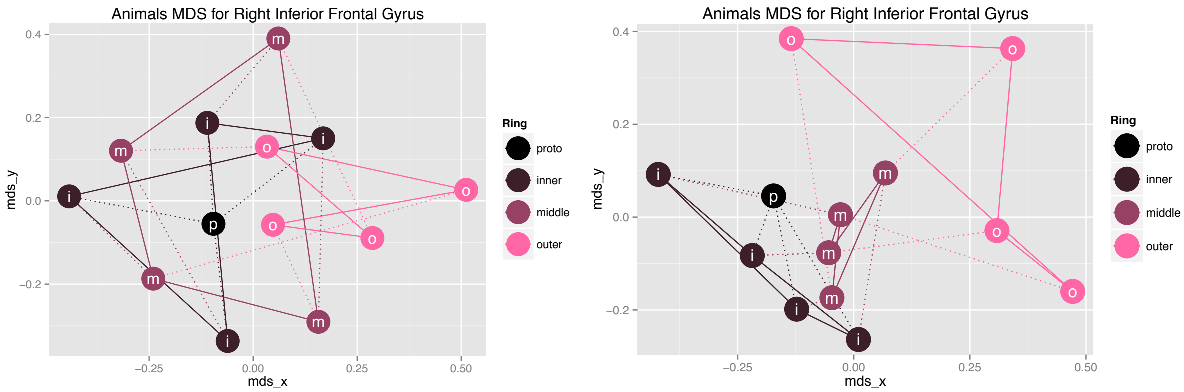

Figure 4: The structure of neural similarity space in the right Inferior Frontal Gyrus while performing the categorisation task. The different rings of animals in the shape-space are coloured light purple, dark purple and black, respectively. **Left:** before training. Note how the different rings of the shape space are highly entangled with each other **Right:** after training. The new post-training neural representations have now disentangled the different rings of the shape-space.

bridge that we are aware of between a computationally transparent approach in machine learning and the neural representational structures that are actually used by the brain.

2.4 Finding low-dimensional representations in the brain: manifolds

The core hypothesis guiding our lab's work is that structure in neural representations reflects structure in the world. A crucial aspect of structure in the world is that a single, simple, underlying cause can give rise to a large number of different sensory experiences. For example, a single object, when viewed from different angles, creates many different projections on the retina. One of the brain's central tasks is to recover that underlying simple cause from the large and seemingly diverse variety of sensory impressions.

Another way of stating this problem, in more mathematical terms, is that the brain needs to discover the low-dimensional cause that underlies the higher-dimensional sensory signal. In machine learning and computer vision, a low dimensional space embedded inside a higher dimensional one is known as a manifold.

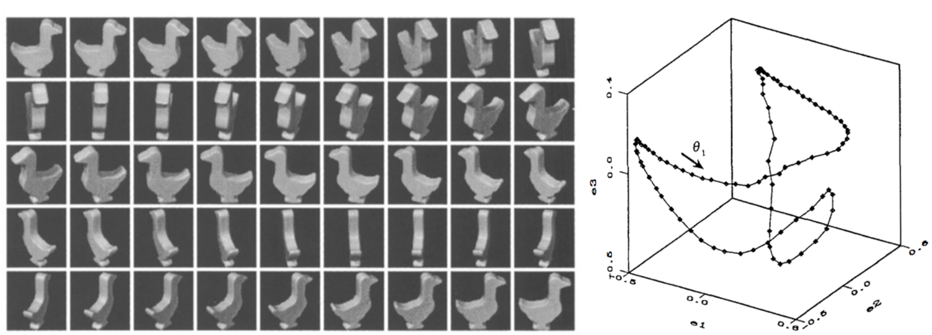

Figure 5: **Left:** different viewpoints of a single 3D duck shape. **Right:** The low-dimensional manifold underlying the different viewpoint images. From Murase and Nayar (1995)

In essence, a manifold is a surface. Consider a piece of paper: it is intrinsically two-dimensional, and continues to be so even if it is scrunched up into a ball. The scrunched up paper is now tangled up throughout a piece of 3D space, but the simpler two-dimensional piece of paper is still there, and constitutes the truest description of the underlying structure.

In computer vision and machine learning, there has been a great deal of interest in finding low-dimensional manifolds underlying visual appearance. The seminal paper was that of Murase and Nayar (1995), who investigated the manifold underlying different viewpoints of a 3D shape.

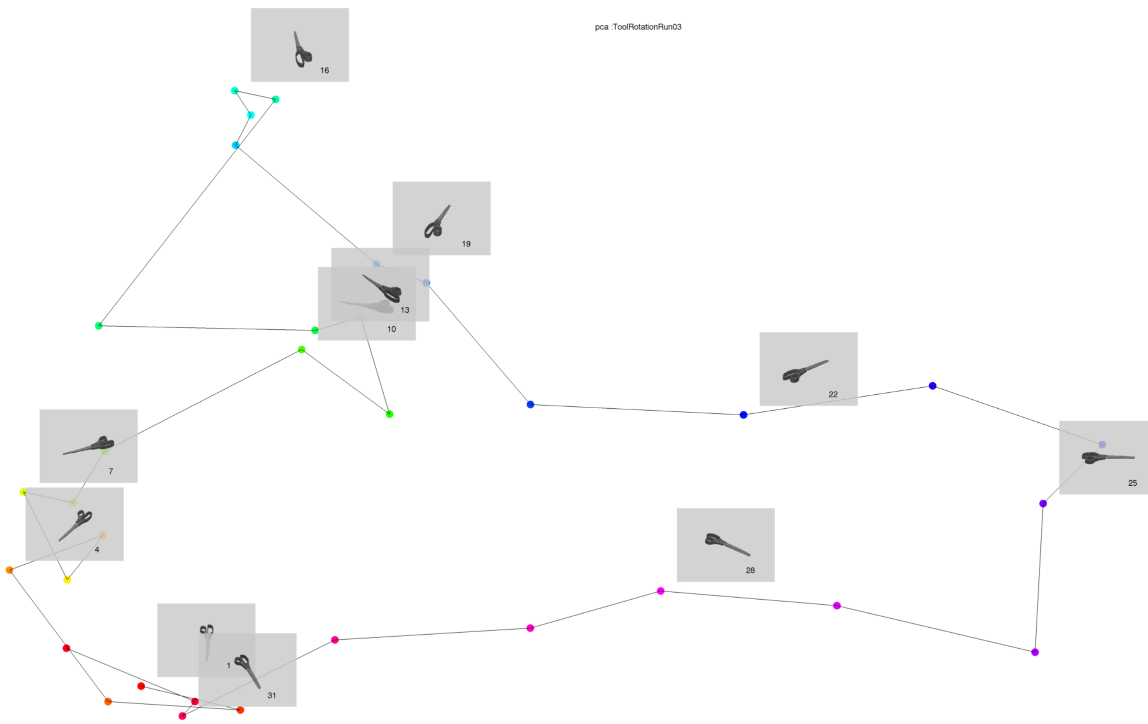

Figure 6: A low-dimensional projection of the pattern-space spanned by fMRI representations of different viewpoints of a rotating tool, in mid-level visual cortex. The colour of the dots represents viewpoint angle of the visually presented 3D shape.

In collaboration with BCS faculty member Brad Mahon, two graduate students in my lab (Xixi Wang, from the Biomedical Engineering department, and Carol Jew from BCS) are investigating whether we can find such manifolds directly in people's neural representational structures. Following the lead of work in computer vision, subjects were presented with videos of slowly rotating objects (tools of various sorts, such as hammers, scissors, etc.).

This work is still in its early stages, but initial results are very promising. Figure 6 shows a two-dimensional projection (using Principal Components Analysis, or PCA) of fMRI activation patterns elicited by different viewpoints of a rotating tool, in mid-level visual cortex. The colour of the dots represents viewpoint angle of the visually presented 3D shape, and the actually presented visual image of the tool is shown next to some of the dots (but not next to all of them, to avoid cluttering the figure). The most notable aspect of the figure is the smooth continuity of the dot-colours, showing that the two-dimensional manifold captures the continuity of the underlying shape rotation.

The promising results form these initial analyses open up many avenues of future exploration. Our long-term goal is to form tight links between work on manifolds in computer vision and empirical findings from the human brain. Artificial machine vision and biological human vision are faced by essentially the same set of computational problems. If we can show that they may solve these problems using similar low-dimensional representations, then an important bridge will be built between two currently disparate lines of research.

2.5 Additional collaborations on vision-related work

In addition to the work described above, the lab has started a new collaboration with BCS faculty member Dick Aslin, spearheaded by my postdoc Ben Zinszer. In this work, we are applying novel neural decoding methods, devised in our lab using fMRI data, to decode functional Near Infrared Spectroscopy (fNIRS) data obtained from infants. Although neural decoding of adult fMRI is now a rapidly growing field of study, its use in infant data has barely begun. We are therefore very excited about the possibilities of extending our work into the important area of developmental cognitive neuroscience. A paper presenting the results of this collaboration is currently under review (Emberson et al., 2015). This work has been using visual stimuli (faces), auditory stimuli (spoken words) and also audio-visual crossmodal stimuli.

2.6 Future directions

The research outlined above opens up many new questions to be explored. The shared theme underlying these questions is that of relating algorithms in computer vision and machine learning to the neural representational structures found in the brain. For example, in kernel learning a key question is how different types of information, e.g. visual and semantic, can be combined; the same question arises for the brain. Similarly, in kernel matrix decomposition, artificial algorithms seek to find lower dimensional representations of stimulus and task space. The brain is confronted with this same task in its need to represent the visual world. Might the solutions reached by artificial and natural systems be convergent?

In the domain of manifold learning of visual representations, one particularly intriguing direction for future exploration is its relation to a set of models from computational neuroscience known as "slow feature analysis", which have been used to model how the brain might learn visual invariances in temporally continuous stimuli (Sprekeler, 2011; Kompella et al., 2012; Wiskott and Sejnowski, 2002), based on the insight that invariances are the aspects of the world which change most continuously from moment to moment. Our initial fMRI study described above, with its use of slow and continuous visual stimuli, is the first to be able to touch on these new interconnections. How the human brain solves the computational challenge of vision is a topic which will provide stimulating research questions for a long time to come.

References

Edelman, S. (1995). Representation of similarity in three-dimensional object discrimination. *Neural Computation*, 7(2):408–423.

Emberson, L., Zinszer, B., Raizada, R., and Aslin, R. (2015). Decoding the infant mind: Multichannel pattern analysis (MCPA) using fNIRS. Under review.

Kleinschmidt, D., Emberson, L., Doko, D., Edelman, and Raizada, R. (2016). How the brain transforms its representational space to learn to solve a linearly inseparable task. In preparation.

Kompella, V. R., Luciw, M., and Schmidhuber, J. (2012). Incremental slow feature analysis: adaptive low-complexity slow feature updating from high-dimensional input streams. *Neural Comput*, 24(11):2994–3024.

Murase, H. and Nayar, S. K. (1995). Visual learning and recognition of 3-d objects from appearance. *International journal of computer vision*, 14(1):5–24.

Schölkopf, B. and Smola, A. J. (2002). *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT press.

Shahbazi, R., Raizada, R., and Edelman, S. (2016). Similarity, kernels, and the fundamental constraints on cognition. *Journal of Mathematical Psychology*, 70:21–34.

Sprekeler, H. (2011). On the relation of slow feature analysis and laplacian eigenmaps. *Neural Comput*, 23(12):3287–302.

Wiskott, L. and Sejnowski, T. J. (2002). Slow feature analysis: unsupervised learning of invariances. *Neural Comput*, 14(4):715–70.