Rajeev Raizada: Statement of research interests

An ongoing research program: computation, brain, and behaviour

My research seeks to uncover how the human brain's neural representations underlie task performance: how suitably-structured representations allow a task to be performed well, and how poorly-structured representations may lead to impaired performance or to learning disabilities. To work towards this goal, I have developed and am continuing to explore new computational methods for analysing the multivariate spatial patterns in fMRI data, and for relating these patterns to individual differences in people's behaviour.

The field of pattern-based fMRI analysis (also sometimes referred to as multivoxel pattern analysis, or MVPA) is currently growing at an explosive rate (Raizada & Kriegeskorte, 2010). Three aspects of my own work set it apart. The most central of these is the goal of relating fMRI patterns to behaviour. Most research in this area currently focuses on "decoding," namely reasoning backwards from observed patterns of fMRI activation to infer which experimental condition gave rise to them. In contrast, my work goes beyond that by asking whether we can "reason forwards" from distributed brain activation to make inferences about people's behaviour: Raizada et al. (2010a) was the first paper to show a relationship between the pattern-separability of people's neural representations and individual differences in their behavioural ability. The second distinctive aspect is that my computational training in neural networks (Grossberg & Raizada, 2000; Raizada & Grossberg, 2001, 2003) has equipped me with the tools to generate new analysis approaches, rather than being forced to use whatever off-the-shelf tools might be available. The new analyses, described below, which link whole-brain pattern separability to people's behavioural performance are illustrations of this (Raizada et al., 2010b). The third differentiating factor in my work is its bigger picture goal of relating these more theoretical questions to real-world problems, specifically the question of how Cognitive Neuroscience can help to diagnose and remediate learning difficulties and educational disadvantage (Raizada et al., 2008; Raizada & Kishiyama, 2010; Raizada et al., under review).

A longstanding goal of my research has been to try to move beyond fMRI studies which simply show "what lit up," and instead to investigate how specific information processing operations take place in the brain's distributed networks. In Raizada & Poldrack (2007b), the method of adaptation-fMRI was used to study how the overlapping neural representations of different phonemes are contrasted against each other during the categorical perception of speech. This led directly to exploration of new pattern-based fMRI analysis methods, which use tools from machine learning to access similar questions but in a more powerful way. While at Dartmouth, I co-supervised a Ph.D. student on a research project building upon that categorical perception work, resulting in a manuscript currently under the second round of review at the Journal of Neuroscience (Lee et al., 2011). A different study of mine also sought to examine parametric connections between brain and behaviour, but this time by looking at fluctuations in performance at at the trial-by-trial level (Raizada & Poldrack, 2007a). This paper also made a conceptual contribution, pointing out how standard subdivisions of attention (goal-driven, stimulus-driven etc.) fail to capture the dimension of how much or how little of the brain's cognitive resources are allocated in response to an attentional challenge. Pursuing further the question of how fMRI can uncover internal aspects of competence that are not revealed behaviourally measured performance, and building upon my interest in the development of language skills, I performed a study of 5-year-old children (Raizada et al., 2008), and found that the hemispheric asymmetry of activation in Broca's area could serve as a marker of language development which remained informative even after behavioural measures of language ability were partialled out.

Recent developments: the power of similarity-space

It is now commonplace for computationally-oriented fMRI researchers to apply powerful classifier algorithms to neural data. However, a key ingredient has been missing. Number-crunching does not in itself generate any insights about representational structure. What is needed is a way of applying insights about similarity-structure to understanding neural representations.

I have recently developed a novel method of neural decoding, which operates entirely within similarity-space. As I describe below, this approach yields novel insights into two longstanding questions in Cognitive Psychology: what makes different people's representations alike, and how the brain represents the meanings of words. It also opens up some potentially very fruitful new directions of future research.

A new method of neural decoding, and what it tells us about how the brain represents the world

Are there common organising principles of neural representation might be which hold across individuals? If there do exist such principles, and if we understood them, then we would be able to decode the neural activation of a given subject, using only information from other people's brains. However, until now this has not been possible.

At the level of specific neural representations, pattern-recognition algorithms have been used to find the multivoxel neural "fingerprints" elicited by given stimulus conditions. However, just as the literal fingerprints on people's hands are idiosyncratic to individuals, the "neural fingerprints" of representations in their brains tend to be subject-unique. A classifier trained on one person's set of neural fingerprints fails if it tries to decode activation patterns from somebody else.

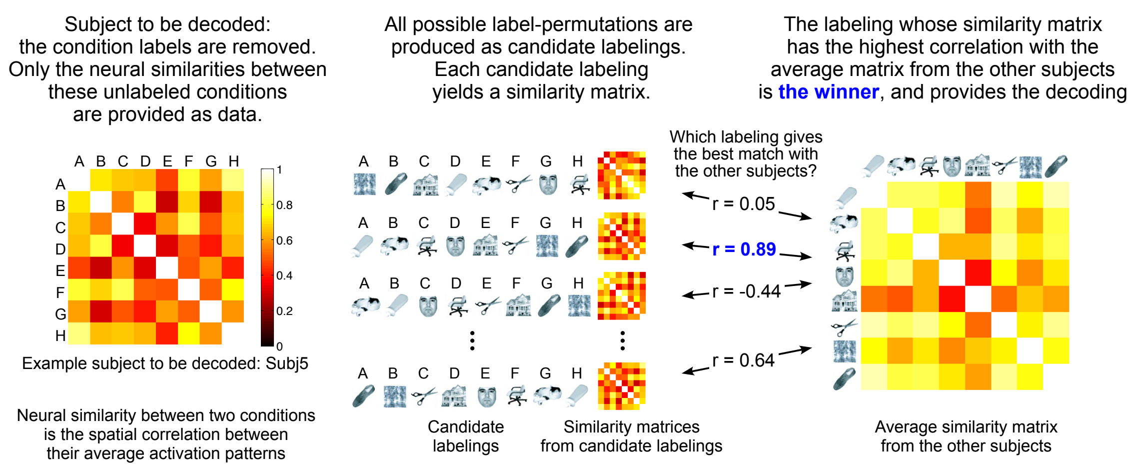

Figure 1: My novel method of across-subject neural decoding, which operates entirely within similarity-space. From Raizada & Connolly (under review).

The key to finding this solution was questioning the seemingly obvious idea that neural decoding should work directly upon neural activation patterns. On the contrary, I have shown that to decode across subjects it is necessary to abstract away from subject-specific patterns of neural activity, and instead to operate on the similarity-relations between those patterns: my new approach performs decoding purely within similarity-space.

My new method is shown in Figure 1. To demonstrate its effectiveness, I used it to perform across-subject decoding on the classic Haxby et al. 2001 dataset of object-elicited activation in ventral temporal (VT) cortex. It is able to decode the neural activation patterns in a given subject's brain using only information from other people's brains, and it achieves an accuracy of 91.7% correct.

A striking aspect of the solution to across-subject decoding presented above is its parallelism to a proposal made more than 20 years ago by the neuro-philosopher Paul Churchland. He proposed, on purely theoretical grounds, that matched structure in people's neural similarity-spaces could explain how different brains can form the same mental representations. He referred to this as "the problem of conceptual similarity across neural diversity." When completely different lines of enquiry, originating respectively from conceptual and empirical concerns, end up converging on the same solution, it is a good indication that they are both being guided by something real.

How does the brain represent the meanings of words?
The question of how the brain represents the meanings of words is one of the most fundamental in human neuroscience, but remarkably little is known about it. There is evidence that certain brain areas such as the left prefrontal cortex are involved in processing word semantics, but that fact in itself does not tell us how those words are actually represented. If we knew how the words were neurally coded, then we would be able to decode those words from patterns of neural activation.

I hypothesised that the brain itself represents words in terms of their semantic similarities with each other, i.e. that neural similarity matches semantic similarity. Similarity-based approaches for representing the meanings of words and concepts have a long history, both in cognitive psychology and in the computational processing of language. However, without a quantitative computational theory linking neural similarity to semantic similarity, the hypothesis that the brain represents meaning in terms of semantic similarity has been neither testable nor even properly expressible. My similarity-based neural decoding method provides precisely such a theory.

Based on this theory, I applied my similarity-based neural decoding method to the data published in Science in 2008 by Tom Mitchell and colleagues. My approach is simple, but achieves substantially more accurate decoding than any previously published result (Raizada, under review).

This is not merely a methodological advance in neural decoding. It tells us something new about how the brain represents meaning. If a simple and theoretically well-grounded approach is able accurately to decode the meanings of words from neural activation, then that suggests that it captures some core properties of the semantic representational structures that are actually used by the brain.

Future research directions: how does the brain generalise?
To survive in the world, our brains need to respond to similar objects in similar ways. However, in achieving that simple-sounding goal, the brain must solve challenging problems, and it does so in ways that are at present only poorly understood. Consider the basic question: "Can I eat this object that I am looking at?" Answering that question requires sensitivity to visual similarity, e.g. recognising that bananas seen from different viewpoints are all the same type of object, and also to conceptual similarity, e.g. understanding that bananas, apples, and nuts are all edible, but that rocks are not.

This ability of the brain to respond flexibly and adaptively to changing circumstances is at the heart of intelligence. Suppose I encounter a fruit on a tree that I have never seen before, and I again ask, "Can I eat this?". A central part of the answer to this question will stem from figuring out how *similar* this is to other things that I know to be edible. In terms of mental representations, this calls for making a task-relevant similarity judgment between a new stimulus and existing representations, thereby *generalising* beyond the realm of prior knowledge. Moreover, if another person comes to the same tree, and also ponders whether or not to eat the fruit, then his or her mental processes will, in crucial respects, have the same structure as mine, even though we are different individuals.

How the brain achieves such generalisation is, at present, only poorly understood. My work opens up new avenues for attacking this longstanding set of problems. In so doing, I will be building upon my training in computational neuroscience (Grossberg & Raizada, 2000; Raizada & Grossberg, 2001, 2003), my experience in applying computational methods to questions in fMRI (Raizada & Poldrack, 2007b; Raizada et al., 2010a,b; Raizada & Kriegeskorte, 2010) and my recent development of completely new neural decoding approaches (Raizada & Connolly, under review; Raizada, under review) Building such a research program is long, hard and often challenging work; I can't think of anything that I'd rather do.

References

Grossberg, S. & Raizada, R. D. S. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research*, *40*(10-12), 1413–1432.

Lee, Y. S., Turkeltaub, P. E., Granger, R. H., & Raizada, R. D. S. (2011). Categorical speech processing in Broca's area: An fMRI study using multivariate pattern-based analysis. *Under second round of review at Journal of Neuroscience*.

Raizada, R. D. S. (under review). Representing the meanings of words: neural similarity matches semantic similarity. *Submitted to PNAS*.

Raizada, R. D. S. & Connolly, A. C. (under review). What makes different people's representations alike: neural similarity-space solves the problem of across-subject fMRI decoding. *Under second round of review at Journal of Cognitive Neuroscience*.

Raizada, R. D. S. & Grossberg, S. (2001). Context-sensitive binding by the laminar circuits of V1 and V2: A unified model of perceptual grouping, attention, and orientation contrast. *Visual Cognition*, *8*(3-5), 431–466.

Raizada, R. D. S. & Grossberg, S. (2003). Towards a theory of the laminar architecture of cerebral cortex: computational clues from the visual system. *Cerebral Cortex*, *13*(1), 100–113.

Raizada, R. D. S. & Kishiyama, M. (2010). Effects of socioeconomic status on brain development, and how cognitive neuroscience may contribute to leveling the playing field. *Frontiers in Human Neuroscience*, *4*, 3.

Raizada, R. D. S. & Kriegeskorte, N. (2010). Pattern-information fMRI: new questions which it opens up, and challenges which face it. *International Journal of Imaging Systems and Technology*, *20*, 31–41.

Raizada, R. D. S., Lebedeva, G. C., Bhagat, S. V., & Kuhl, P. K. (under review). Differences in parental language predict the structure of Broca's area in five-year-old children.

Raizada, R. D. S. & Poldrack, R. A. (2007a). Challenge-driven attention: interacting frontal and brainstem systems. *Frontiers in Human Neuroscience*, *1*, 3.

Raizada, R. D. S. & Poldrack, R. A. (2007b). Selective amplification of stimulus differences during categorical processing of speech. *Neuron*, *56*(4), 726–740.

Raizada, R. D. S., Richards, T. L., Meltzoff, A., & Kuhl, P. K. (2008). Socioeconomic status predicts hemispheric specialisation of the left inferior frontal gyrus in young children. *Neuroimage*, *40*(3), 1392–1401.

Raizada, R. D. S., Tsao, F.-M., Liu, H.-M., Holloway, I. D., Ansari, D., & Kuhl, P. K. (2010b). Linking brain-wide multivoxel activation patterns to behaviour: Examples from language and math. *Neuroimage*, *51*(1), 462–71.

Raizada, R. D. S., Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2010a). Quantifying the adequacy of neural representations for a cross-language phonetic discrimination task: Prediction of individual differences. *Cerebral Cortex*, *20*(1), 1–12.