Turtles all the way down? Psychometric approaches to the reduction problem
Kievit, R.A.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
Appendix C

Glossary for Chapter 6

**Decoding model:** A model that takes brain-activity patterns as input and “predicts” the experimental condition (e.g. the stimulus). Significant decodability indicates that information about the experimental condition is present in the region. A decoder can be used to model readout of a population code and predict the code’s reflection in downstream regions and behaviour.

**Encoding model:** A model that takes the stimulus as input and predicts brain activity. An encoder can be used to model the brain’s processing of the stimulus, which gives rise to a given area’s population-code representation.

**Explicit representation:** A neuronal representation of a stimulus property that allows immediate readout of the property by downstream neurons. If the property can be read out by means of a linear combination of the neurons’ activities or by a radial basis function, the property is explicitly represented. (For example, the category of a visual object is implicitly represented in the retina and explicitly represented in inferior temporal cortex.)

**Geometric model of similarity:** A model of an internal multidimensional representational space originally devised to explain human dissimilarity judgments and patterns of generalization from examples to novel stimuli. By measuring brain-activity patterns, we can attempt to find the neuronal implementation of the internal representational space. In the present context, this body of ideas is important also because we can apply the concepts of similarity and generalization to gain a functional understanding of neuronal population codes and their transformation across stages of processing. The conceptual function of the geometric model for understanding neuronal computation is independent of its original goal, and its successes and limitations, as a cognitive theory of similarity judgments and generalization behaviour.

**Represented information:** The mutual information between stimuli and response patterns. This comprises any differences between stimuli that can in principle be decoded from a neuronal population code. Given limited data, we cannot fit arbitrarily complex decoding models in practice. Thus estimates of the represented information are negatively biased. The data processing inequality states that information can only be lost, never gained, along stages of processing. The retina has all the represented information about visual stimuli that any brain region has.

**Linear readout:** A decoding scheme based on a weighted sum of the input neurons’ activities. Linear decoders are useful because they are realistic to fit given limited brain-activity data and any information they reveal about the stimulus can be thought of as explicitly represented (see above).

**Multidimensional scaling:** A procedure by which we can arrange n items in a d dimensional space, such that their distances in the space best reflect their
dissimilarities (by different metric or nonmetric criteria). The technique can be used, for example, to attempt to recover the internal representational space (which is typically assumed to be Euclidean) from dissimilarity judgments. It is also useful for producing 2-dimensional arrangements that best visualise the distances of the items in a higher dimensional space, such as a neuronal population code. (Note that the MDS objective to best represent the original distances in d dimensions is distinct from the objective of principal components analysis to find the d-dimensional linear subspace that explains maximum variance.)

**Population code:** A scheme for encoding information thought to be important to the organism in the activity of a population of neurons. The perhaps simplest scheme uses only the firing rates of the neurons. More complex codes also use the information contained in the precise temporal patterns of spikes and their relationships between neurons. For each neuron, a “tuning curve” describes how the firing rate reflects particular stimulus properties. A population code for a particular type of information is “good” when it represents the information in a format that can be read out by neurons receiving input from the population. Moreover, a good code should be efficient in the sense of achieving the required precision without wasting neuronal and energy resources.

**Representational dissimilarity matrix (RDM):** A square matrix indexed horizontally and vertically by the stimuli (or experimental conditions) and containing a dissimilarity index in each cell, which compares the two brain-activity patterns associated with the stimuli labelling the row and column. An RDM provides a useful characterization of the representational geometry (see below) for a limited set of experimental stimuli. If the same activity pattern estimates are used for the vertical and horizontal dimensions, the RDM is symmetric about a diagonal of zeros. If independent pattern estimates are used for the vertical and horizontal dimensions, the RDM contains entries comparing independent pattern estimates for identical stimuli (reflecting measurement noise) along its diagonal, and two alternative dissimilarity estimates for each stimulus pair in symmetric off-diagonal positions.

**Pattern component modelling (PCM):** An analysis technique that decomposes the variance of the brain-activity patterns associated with a set of stimuli (or experimental conditions) into components that reflect different predefined factors such as the stimulus category, within-category variance, and measurement noise. PCM can provide useful summary descriptions of the representational geometry.

**Pattern-information analysis:** An approach to the analysis of brain-activity data, in which the activity patterns within functional brain region are analysed multivariately, as a population code. For example, a cortical area may be recorded from invasively with an array of electrodes or imaged with fMRI. The responses measured with electrodes or fMRI are typically also restricted to no more than a few hundred channels per brain region and can be viewed as a sample of the neuronal population activity. The subsampling means that a lot of information is lost in either case. Results should be interpreted as lower bounds on the information actually present in a region. fMRI is additionally compromised by the fact that voxels reflect neuronal activity only indirectly
Psychometric approaches to the reduction problem

through hemodynamics. A voxel’s response to increased neuronal activity is thought to reflect a combination of sub-threshold activity and neuronal firing, averaged locally across space and time.

**Representational geometry:** The geometry of a set of entities represented as points in a space spanned by the dimensions of a neuronal population code. Representational geometry focuses on the relationships between the entities, rather than on single entities, and on geometric properties including distances in the high-dimensional space, rather than on differences in the activity of single neurons. Representational geometry provides a useful intermediate level of description that helps us abstract from idiosyncrasies of individual brains and highlights the representational properties that are key to the computational goals of the brain.

**Similarity:** A subjective notion of the relationship between two objects that reflects the degree to which an organism distinguishes them. Similarity is inherently relative, because any two non-identical objects will share some properties and differ in other properties. Judging similarity from object properties necessarily implies some choice and weighting of the properties. In the interest of its survival and reproduction, an organism should regard two objects as similar, if they require the same behavioural response, for example, if one can replace the other (positive objects), or if both pose the same danger (negative objects). Which properties are relevant and irrelevant and how they should be weighted depends on the individual, its current goals, and the context. We can apply the concept of similarity not only at the level of the entire organism but also at the level of individual brain representations. This enables us to view neuronal computation as the stage-by-stage transformation of population-code representational similarity, whose function is to be understood in the larger context of the goals of brain information processing.

**Single-neuron computational model:** A model that predicts the responses of a particular neuron to novel stimuli. For example, responses of primary visual simple and complex cells can be modelled using localized linear filters of the input image and simple nonlinearities. For higher-level brain regions like IT, it is currently impossible to fit single-neuron computational models. The problem is not only that the model space is insufficiently defined, but that the number of parameters to be fitted (for any plausible model space) is too large given the amount of stimulus-response data that can be acquired.