CellPress

Opinion

Painful Issues in Pain Prediction

Li Hu[1,2,3,*] and Gian Domenico Iannetti[2,*]

How perception of pain emerges from neural activity is largely unknown. Identifying a neural 'pain signature' and deriving a way to predict perceived pain from brain activity would have enormous basic and clinical implications. Researchers are increasingly turning to functional brain imaging, often applying machine-learning algorithms to infer that pain perception occurred. Yet, such sophisticated analyses are fraught with interpretive difficulties. Here, we highlight some common and troublesome problems in the literature, and suggest methods to ensure researchers draw accurate conclusions from their results. Since functional brain imaging is increasingly finding practical applications with real-world consequences, it is critical to interpret brain scans accurately, because decisions based on neural data will only be as good as the science behind them.

Trends

Predicting perceived pain from brain activity has enormous implications: 'pain signatures' from brain imaging data are increasingly used as evidence for pain perception in minimally conscious patients or infants, or in legal settings.

Sophisticated machine-learning algorithms are increasingly applied to functional brain-imaging data with two main objectives: (i) identifying a specific neural 'pain signature'; and (ii) predicting perceived pain from brain activity.

While machine-learning approaches hold considerable promise for pain research, they are fraught with interpretive difficulties: disregarding the tight match between machine-learning protocol design and the desired study objectives could lead to incorrect interpretation of results.

Machine Learning in Pain Research: Objectives and Protocols

Pain, as any other conscious sensation, is determined by a specific pattern of neural activity at the cortical level [1,2]. To understand the perception of pain, many researchers use non-invasive functional neuroimaging techniques [3,4], such as electroencephalography (EEG), magnetoencephalography (MEG), positron emission tomography (PET), and, especially, functional magnetic resonance imaging (fMRI). With these tools, researchers can now attempt to achieve the following key objectives: (i) identify temporal and spatial patterns of neural activity that could serve as a cortical signature for human pain perception [5–8]; and (ii) establish whether these patterns, or any other physiological measures of brain activity, can be used to reliably predict perceived pain [7,9–14]. Achieving these objectives, which would have dramatic basic and clinical implications, is increasingly attempted through the application of sophisticated **machine-learning** (see Glossary) algorithms to interpret functional brain-imaging data [15–18]. However, correct interpretation requires proper protocol design and careful inferences. Here, we highlight some of the pitfalls of applying machine-learning techniques to functional brain-imaging data related to pain perception, especially in light of recent divergent conclusions in the literature, and suggest possible remedies.

Machine learning is a scientific discipline exploiting algorithms that can learn and make **predictions** from data [19–21]. When applied to functional brain-imaging data, machine learning has the potential to: (i) identify response features that specifically encode a given experimental variable (e.g., the categories of visual objects [22]); and (ii) decode measured data to predict subjective percepts and intentions (e.g., the pain intensity reported by an individual [9]) (Box 1). Therefore, it is not surprising that machine learning has received immense interest in systems neuroscience, and it is now increasingly used in the field of human pain [7,9–14,23,24].

While machine-learning techniques hold considerable promise for pain research, investigators must take special care to match machine-learning protocol design to the desired study

[1]Institute of Psychology, Chinese Academy of Sciences, Beijing, China
[2]Department of Neuroscience, Physiology and Pharmacology, University College London, London, UK
[3]Key Laboratory of Cognition and Personality (Ministry of Education) and Faculty of Psychology, Southwest University, Chongqing, China

*Correspondence: hulitju@gmail.com (L. Hu) and g.iannetti@ucl.ac.uk (G.D. Iannetti).

 http://dx.doi.org/10.1016/j.tins.2016.01.004

Box 1. Encoding, Decoding, and Reverse Inference

In functional brain imaging, 'encoding' refers to the identification of a statistical dependency between experimental variables (e.g., pain perception) and measured brain responses. This encoding procedure is normally achieved using the traditional voxel-by-voxel mass-univariate analysis of fMRI time series (using, for example, general linear modeling: GLM, Figure I).

In contrast, 'decoding' comprises predicting the same experimental variables based on the measured brain responses. This decoding procedure is typically achieved using machine learning (e.g., multivoxel pattern analysis, MVPA, Figure I), which is based on certain features of the fMRI response (e.g., patterns of fMRI activity distributed over many voxels).

Reverse inferences are logically flawed deductions based on affirming the consequent (e.g., if A determines B, when B is observed one infers that A has occurred). Reverse inferences are notoriously frequent in functional neuroimaging research, and typically consist in inferring a particular experimental variable (e.g., the perception of pain) from a given pattern of brain activation (e.g., the so-called 'pain matrix') [37,38]. Notably, reverse inferences have a probability of being correct, which depends on the exclusivity of the relation between the experimental variable and the recorded response (i.e., it depends on how many variables other than A determine B).

Even if decoding is the reverse prediction of experimental variables from the measured brain response, decoding is conceptually different from reverse inference: indeed, in most practical applications, decoding analysis does not require that the relation between the experimental variable and the corresponding brain response is exclusive. For example, most currently available pain prediction algorithms rely on features of the brain response that are not tested for their necessity or sufficiency for the occurrence of pain perception.

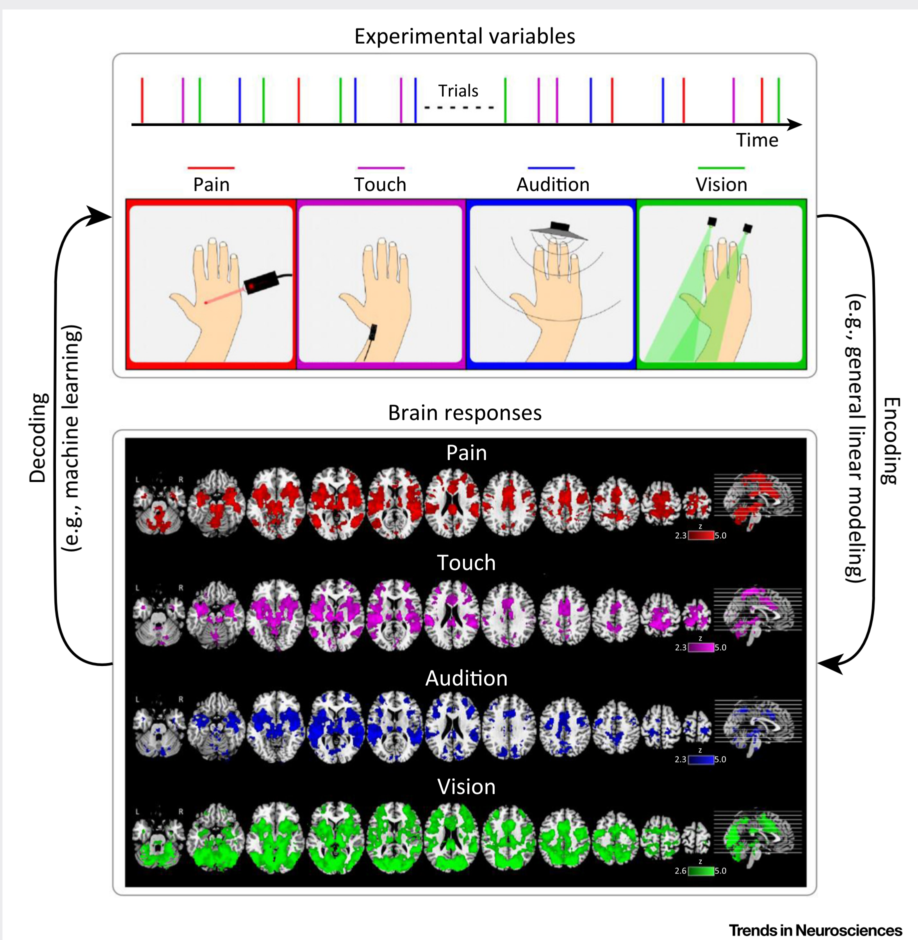

Figure I. Relationship between Encoding (Identifying the Statistical Dependency between Experimental Variables and Brain Responses) and Decoding (Predicting Unknown Experimental Variables from the Brain Responses). Bottom panel modified from [29].

Glossary

Machine-learning prediction: once machine learning has identified a response pattern associated with an experimental variable, it can be used to predict that experimental variable on the basis of the detected response pattern.

Machine learning: an analysis approach that comprises using the ability of computers to learn from, and make predictions from, different kinds of data. When applied to functional brain images, machine learning can be used to detect response patterns (e.g., intensity and spatial distribution of fMRI signals) associated with a given experimental variable (e.g., the intensity of pain perception).

Multivoxel pattern analysis (MVPA): a kind of machine-learning technique that identifies condition-specific spatial patterns of fMRI responses distributed across different voxels. These patterns of activity can be used to predict the occurrence of different experimental variables (e.g., different levels of subjective pain, or pain vs touch).

Neural signature: a feature of the brain response that is uniquely associated with a given experimental variable. To identify conclusively a neural signature, it is crucial to ensure that its relation with the experimental variable is exclusive (i.e., that other experimental variables do not produce the same pattern of brain response).

Pain prediction: the process of estimating unknown subjective intensity of pain perception using experimentally measured functional brain-imaging data. True pain prediction must not use prior knowledge about subjective reports of pain intensity when testing the prediction performance.

Prior knowledge: in the context of machine learning, refers to the information about the experimental variables that, although available, should not be used when testing the performance of the machine-learning classifier in predicting an experimental variable. The incorporation of prior knowledge into the training is a necessary aspect of machine learning. By contrast, exploiting prior knowledge when testing the algorithm performance is incorrect, and results in an artificial inflation of performance (false positive results).