OXFORD

Social Cognitive and Affective Neuroscience, 2021, 838–848

doi: https://doi.org/10.1093/scan/nsaa063
Advance Access Publication Date: 5 May 2020
Original Manuscript

Special Issue: Computational Methods in Social Neuroscience

Conducting decoded neurofeedback studies

Vincent Taschereau-Dumouchel,[1,2] Aurelio Cortese,[1] Hakwan Lau,[1,2,3,5,6] and Mitsuo Kawato[1,4]

[1]Department of Decoded Neurofeedback, ATR Computational Neuroscience Laboratories, Kyoto 619-0288, Japan, [2]Department of Psychology, UCLA, Los Angeles, CA 90095, USA, [3]State Key Laboratory of Brain and Cognitive Sciences, University of Hong Kong, Hong Kong, [4]RIKEN Center for Advanced Intelligence Project, ATR Institute International, Kyoto, Japan, [5]Brain Research Institute, UCLA, Los Angeles, CA 90095, USA and [6]Department of Psychology, University of Hong Kong, Pokfulam, Hong Kong

Correspondence should be addressed to Vincent Taschereau-Dumouchel, Department of Decoded Neurofeedback, ATR Computational Neuroscience Laboratories, Kyoto 619-0288, Japan. E-mail: vincenttd@ucla.edu.

Abstract

Closed-loop neurofeedback has sparked great interest since its inception in the late 1960s. However, the field has historically faced various methodological challenges. Decoded fMRI neurofeedback may provide solutions to some of these problems. Notably, thanks to the recent advancements of machine learning approaches, it is now possible to target unconscious occurrences of specific multivoxel representations. In this tools of the trade paper, we discuss how to implement these interventions in rigorous double-blind placebo-controlled experiments. We aim to provide a step-by-step guide to address some of the most common methodological and analytical considerations. We also discuss tools that can be used to facilitate the implementation of new experiments. We hope that this will encourage more researchers to try out this powerful new intervention method.

Key words: decoded neurofeedback; multivoxel pattern analysis; real-time functional magnetic resonance imaging

Introduction

Thanks in part to the development of sophisticated machine learning techniques (Cohen *et al.*, 2017), scientists can now read out specific brain representations using multixovel pattern analysis (Haxby *et al.*, 2001; Kriegeskorte *et al.*, 2008a; Haxby *et al.*, 2014; Taschereau-Dumouchel *et al.*, 2019). However, we rarely think of these methods as tools for intervention. Here, we suggest that when coupled with associative learning by pairing multivoxel patterns with online reward, punishment or another stimulus, we can exploit the power of machine learning for the manipulation of brain activity.

Decoded neurofeedback is a special type of fMRI-based multivoxel neural reinforcement. It aims to provide participants with control over some specific brain processes. Essentially, in this method, a terminal monetary reward is paired with feedback representing the activation likelihood of a targeted multivoxel pattern (see Figure 1; Watanabe *et al.*, 2018). Importantly, participants are not provided with any explicit training strategy; they are simply asked to manipulate their brain activity in order to maximize their reward. As a result, both the participants and the experimenters can be kept unaware of the precise nature of the intervention. Importantly, this allows one to conduct double-blind placebo-controlled interventions, which represent a high level of experimental rigour that few psychological intervention methods can achieve (see (deBettencourt *et al.*, 2015; Young *et al.*, 2017; Taschereau-Dumouchel *et al.*, 2018a, 2018b; Bu *et al.*, 2019).

Received: 22 August 2019; **Revised:** 13 February 2020; **Accepted:** 27 April 2020

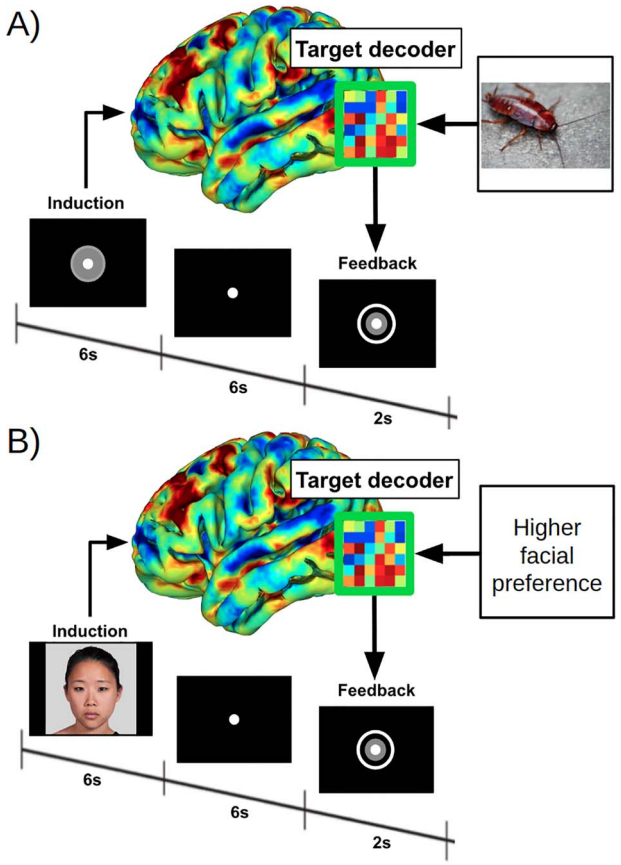


Fig. 1. Sequences of events in one trial of (A) decoded neurofeedback and (B) associative decoded neurofeedback. Echo-planar images are acquired in the fMRI scanner during the induction period. Basic preprocessing is conducted online before the data is inputted into the target decoder. This target decoder was previously trained offline to provide an activation likelihood of the target category (e.g. 70% probability of the brain representing a cockroach). Decoded neurofeedback involves pairing this activation likelihood with a reward and providing feedback to the participant that represents this information (e.g. the diameter of the inner gray circle). Two methods have been used previously. (A) Pairing the activation likelihood with a reward (see Shibata *et al.*, 2011; Koizumi *et al.*, 2016; Cortese *et al.*, 2016b; Taschereau-Dumouchel *et al.*, 2018b). (B) Presenting a stimulus visually during the induction period, which might allow the creation of a new unconscious association between the visual stimuli and the target decoder (see Amano *et al.*, 2016; Shibata *et al.*, 2016). (The face stimulus was adapted from Strohminger *et al.* (2016)).

Two main approaches of decoded neurofeedback have been used. The first approach involves directly pairing a monetary reward with the activation of a specific multivariate pattern (see Figure 1A). The second approach is what we can call associative decoded neurofeedback (Amano *et al.*, 2016). It has been used to create an unconscious association between a specific multivariate pattern and a target stimulus presented visually during reinforcement (see Figure 1B).

Using these approaches, previous studies showed that it is possible to target specific processes, such as perceptual learning (Shibata *et al.*, 2011; Amano *et al.*, 2016), metacognition (Cortese *et al.*, 2016a), emotion perception (Shibata *et al.*, 2016), finger movements (Oblak *et al.*, 2020) and physiological threat reactivity (Koizumi *et al.*, 2016; Taschereau-Dumouchel *et al.*, 2018b, 2020), directly and unconsciously in the human brain. Importantly, these interventions changed specific behavioral or physiological outcomes related to the targeted representations. This provides strong evidence suggesting a causal link between the targeted brain representations and the associated outcomes. This represents an important feature of decoded neurofeedback because, in human neuroscience, few tools can help establish such a causal link at the level of brain representations with high specificity.

Our goal is to provide an overview of decoded neurofeedback research with a specific focus on the technical considerations for the implementation of new research protocols. We will cover the history and scientific principles of decoded neurofeedback before discussing the step-by-step procedure to develop new decoded neurofeedback experiments. While decoded neurofeedback has also been successfully conducted in electroencephalography (EEG) see (Bu *et al.*, 2019; Faller *et al.*, 2019; Keynan *et al.*, 2019), here we will primarily focus on fMRI applications.

How is decoded neurofeedback different from traditional neurofeedback approaches?

The first successful neurofeedback experiments were conducted in the late 1960s and early 1970s. Early studies showed that primates could be trained to increase the firing rate of specific precentral neurons (Fetz, 1969). Around the same period, experiments using human EEG also suggested that participants could be trained to regulate the power of some specific frequency bands (Kamiya, 1968, 2011; Beatty *et al.*, 1974). These first demonstrations sparked an interest in the method, and many investigations followed using diverse approaches such as fMRI (Weiskopf, 2012), EEG (Evans and Abarbanel, 1999), electrocorticography (Gharabaghi *et al.*, 2014), and functional near-infrared spectroscopy (Kober *et al.*, 2014). The potential for therapeutic applications was quickly identified (Hardt and Kamiya, 1978), and numerous studies explored the possibility of conducting treatments using neurofeedback (Hammond, 2005; Arns *et al.*, 2009; Cortese *et al.*, 2016b; Mennella *et al.*, 2017; Young *et al.*, 2017; Van Doren *et al.*, 2019).

However, concerns were also raised regarding the scientific validity of these interventions (Thibault *et al.*, 2015; Schabus *et al.*, 2017; Schönenberg *et al.*, 2017). Critics notably pointed out that many neurofeedback interventions lacked specificity and were prone to the placebo effect. Decoded neurofeedback emerged from this scientific tradition and uses modern solutions to address problems previously raised in the field. In the next sections, we discuss the technological and methodological improvements that differentiate decoded neurofeedback from traditional neurofeedback experiments.

One important feature of decoded neurofeedback is the reliance on modern machine learning approaches (Cohen *et al.*, 2017). Many previous neurofeedback studies determined the intervention targets using univariate group statistics. This can be problematic because such targets are known to vary substantially between individuals (Gonçalves *et al.*, 2006). As such, it is difficult to determine if the provided feedback accurately reflects the intended brain process. For this reason, decoded neurofeedback is typically preceded by an initial experimental session called the decoder construction session. During this session, a multivariate predictive model (also called a decoder) is trained to predict a specific brain process. Importantly, the accuracy of this prediction can be determined within-subject and at the single-trial level. For instance, in order to design an experiment to train selective attention, we first need to acquire brain data when the participant is attending or not attending to visual stimuli. Using machine learning

algorithms, it will then be possible to train a decoder of selective attention and to determine how accurately this decoder can predict the data of the target participant. This process allows one to determine if the targeted decoder is indeed accurate for each participant. Previous decoded neurofeedback experiments were primarily conducted using binary decoders in a classification framework. Using this approach, the feedback provided to the participant represents the activation likelihood of the decoder. However, it is also possible to conduct decoded neurofeedback in a regression framework with a continuous outcome. Here, the feedback represents the predicted value of the continuous outcome. For instance, a previous study provided feedback using the continuous value predicted by a decoder of facial preference (Shibata *et al.*, 2016). As such, decoded neurofeedback experiments can be designed to target both dichotomous and continuous constructs.

Decoded neurofeedback is also conducted without providing any explicit training strategies. Some previous neurofeedback approaches (e.g. de Charms *et al.*, 2005) provided such explicit strategies to participants in order to help facilitate their learning (e.g. 'learn to enhance your control over a localized brain region associated with attention'). However, this approach makes it difficult to dissociate the effect of neurofeedback from the effect of the strategy (de Charms *et al.*, 2005). This is important because if the explicit strategy is the main driver of the observed effect, neurofeedback may actually be irrelevant, and conducting only a cognitive intervention would be sufficient. For this reason, in decoded neurofeedback, participants are provided with as little information as possible regarding the task. They are simply asked to maximize their reward by modulating their brain activity (see Figure 1).

In decoded neurofeedback, it is also important to determine the level of awareness associated with the intervention. This is important because even if no explicit strategies are provided, it might still be possible for the participant to guess the purpose of the intervention and to use a related mental strategy. In order to document this possible confound, experimenters usually explicitly ask participants to report the strategies used during the procedure. In previous decoded neurofeedback studies, participants typically reported using various unrelated strategies, which suggests that they were unaware of the nature of the intervention (Shibata *et al.*, 2011; Koizumi *et al.*, 2016; Cortese *et al.*, 2016a; Taschereau-Dumouchel *et al.*, 2018a). This approach can be complemented by explicitly asking participants, at the end of the experiment, what was the purpose of the intervention using a two or three forced-choice question (e.g. do you think the purpose of the intervention was X or Y?). If the participants were unaware of the nature of the task, we can expect them to answer randomly.

Another key aspect of decoded neurofeedback is the reliance on an external monetary reward provided using intermittent feedback. These two reinforcement parameters were explored in previous studies and present some clear advantages for decoded neurofeedback. First, it was shown that neurofeedback is more effective when paired with a monetary reward, as opposed to visual feedback only (Sepulveda *et al.*, 2016). Second, recent computational modeling suggested that intermittent feedback, as opposed to continuous feedback, is preferable when conducting implicit neurofeedback training (Oblak *et al.*, 2017). One critical advantage of intermittent feedback is that the induction period is clearly bounded in time (see Figure 1). As such, participants can associate the reward with their brain activity during a precise time period (Watanabe *et al.*, 2017). This is typically difficult to achieve in fMRI when the feedback is provided continuously (however, see Ramot *et al.*, 2017; Ramot and Gonzalez-Castillo, 2019).

Why use decoded neurofeedback?

Decoded neurofeedback can reveal the causal link between brain representations and their associated behavioral or psychological outcomes. For instance, in a previous experiment, the representation of a Gabor patch in the primary visual cortex was targeted using decoded neurofeedback (Shibata *et al.*, 2011). This training was associated with perceptual learning specific to the orientation of the targeted representation. Such a precise behavioral consequence was achieved while (i) the target of the intervention remained unknown to participants, (ii) no related physical stimuli were presented and (iii) little information was transmitted to brain regions outside of the targeted area. Taken together, these represent strong evidence for the existence of a causal link between the targeted brain representation and the behavioral outcome. As such, decoded neurofeedback can reveal causal links where most other neuroimaging tools can only reveal associations. Furthermore, decoded neurofeedback represents a rather unique opportunity to test the functional consequences of forming unconscious associations in the brain. This highlights the potential importance of decoded neurofeedback for human neuroscience, as few non-invasive methods can rigorously assess such causal hypotheses.

The mechanisms by which decoded neurofeedback might achieve such an influence are still incompletely understood. Three main mechanisms of action have been suggested. First, the training might lead to an increased activation likelihood of the target decoder. This has been observed in some previous decoded neurofeedback experiments (Shibata *et al.*, 2011; Amano *et al.*, 2016; Taschereau-Dumouchel *et al.*, 2018b). One possibility is that the brain might be capable of using visual feedback in order to more reliably activate the target representation. This might be achieved in a process akin to operant conditioning (Bray *et al.*, 2007). While this task appears to be difficult to solve, recent results indicate ways in which this could be achieved. For instance, Shibata *et al.* (2019) recently showed that fMRI brain signals during decoded neurofeedback training are drastically less complex than previously thought. Only a limited number of components were necessary to explain most of the induction data, which suggests that the space of possible multivoxel patterns to explore could be in fact small. As such, conducting operant conditioning within a limited number of trials might be feasible (Cortese *et al.*, 2019). This possibility is supported by computational simulations conducted with similar parameters as those used in previous decoded neurofeedback experiments. These simulations showed that an increased activation likelihood follows the reinforcement of a multivariate pattern in an array of simulated neural activity (Oblak *et al.*, 2017; Shibata *et al.*, 2019). Therefore, decoded neurofeedback training might be achieved through a simple reinforcement learning rule.

Second, during decoded neurofeedback, participants might also learn to unconsciously associate the multivoxel pattern with a reward, which might potentially act as a form of counter-conditioning. This process is closely related to the mechanism of exposure-based psychotherapy, which is one of the most effective psychological treatments for anxiety disorders (Craske *et al.*, 2008). Through inhibitory learning (Craske *et al.*, 2008), exposure to a feared object allows the development of a new association between the feared object and a more neutral affect. Previous experiments indicated that decoded neurofeedback might achieve a similar process. For instance, Koizumi *et al.*

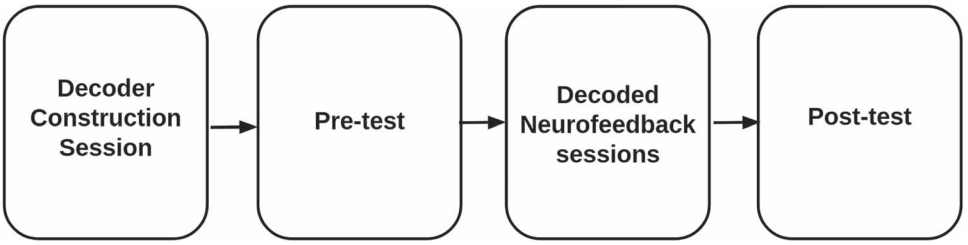


Fig. 2. General design of a decoded neurofeedback intervention. In order to conduct decoded neurofeedback interventions, we first need to conduct a decoder construction session. This session will notably allow us to determine an accurate decoder to be used for the intervention. Typically, the decoded neurofeedback sessions will be preceded and followed by experimental sessions that can allow us to determine if the intervention successfully changed the targeted process (here, pre- and post-test).

(2016) showed that decoded neurofeedback could decrease the physiological threat reactivity (i.e. as indexed by amygdala and skin conductance reactivity) associated with conditioned stimuli. Furthermore, pairing a reward with the unconscious activation of the multivoxel pattern of a feared animal led to a similar decrease in physiological threat reactivity (Taschereau–Dumouchel *et al.*, 2018a). In addition, Chiba *et al.* (2019) recently suggested that a similar effect could be achieved with patients presenting post-traumatic stress disorders. Although the sample size was limited, this study still provides the first evidence that decoded neurofeedback might be effective at decreasing the symptoms of patients presenting anxiety disorders. While the mechanisms behind the effect of these interventions are still uncertain and need further elucidation, it is possible that both counterconditioning and an effect akin to exposure (as a result of an increased activation likelihood) might have occurred (see Chiba *et al.*, 2019 for meta analyses on these two effects).

Lastly, another interesting feature of decoded neurofeedback is the possibility of creating new unconscious associations between a stimulus presented visually (or through other sensory channels) and a targeted multivoxel pattern (Figure 1B; Amano *et al.*, 2016; Shibata *et al.*, 2016). Of particular interest, a study showed that the preference for a previously neutral face can be changed as a function of the targeted multivoxel pattern. More precisely, targeting the pattern of a face with a higher preference led to a greater preference of the neutral faces after decoded neurofeedback, while targeting the pattern for lower preference had the opposite effect (Shibata *et al.*, 2016).

How to perform decoded neurofeedback?

Conducting decoded neurofeedback presents technical challenges. The next sections are meant to detail the general methodological steps of decoded neurofeedback (Figure 2) and discuss some of the main challenges. This information is meant for the novice reader who would like to get acquainted with the implementation of decoded neurofeedback.

Multiple resources related to decoded neurofeedback can be found at https://bicr.atr.jp/decnefpro/. This includes access to scripts to conduct both decoded neurofeedback and functional connectivity neurofeedback. Furthermore, functional brain scans in nifti format, anonymized anatomical scans, decoder information, demographic data and experimental conditions of previous decoded neurofeedback studies are similarly available. These resources are available for non-commercial uses and upon the reception of a signed agreement on terms and conditions for academic use.

General technical and methodological considerations

Some may expect decoded neurofeedback to require particularly powerful computers to be conducted. However, while decoder construction sometimes requires great computational power, the actual reinforcement sessions can be conducted with relatively standard laptop computers (e.g. Intel Core i7 processors with 16 GB of RAM memory). Using such hardware, most online analyses can be implemented within a few hundred milliseconds. As such, conducting decoded neurofeedback may not be much more computationally demanding than conducting ROI-based feedback. The most important hardware consideration for the proper implementation of a real-time procedure is the speed of the connection between the scanner and the computer performing the online analyses. A good approach is to set up a wired local network through optic fibers to directly transfer the DICOM files to a 'shared folder' on the hard drive of the processing computer. This way, the online processing of the acquired images can start as soon as they are made available by the fMRI scanner.

Other general considerations pertain to the design of the interventions. How many sessions of how many runs should be provided? How many participants should be included? There is no straightforward answer to these questions. However, the state of current knowledge has been summarized in a recent paper that can be consulted when designing new experiments (Watanabe *et al.*, 2017). Some useful computational approaches have also been discussed to determine some of the key experimental parameters (Oblak *et al.*, 2019; Ramot and Gonzalez-Castillo, 2019). Most of the previous protocols used runs of 16 trials in such a way that roughly 10 runs could be conducted in each session. The number of sessions has varied among studies; some included 3 sessions carried out on 3 consecutive days, while others included as many as 10 sessions. The number of participants has also varied; some studies included 10 participants (Shibata *et al.*, 2011; Cortese *et al.*, 2016b) while others up to 24 (Shibata *et al.*, 2016). If a similar study was previously conducted, we advise to determine the sample size using the effect size previously reported as well as the previous experimental parameters. Information regarding the effect sizes of previous projects can be found here in Table 5: https://bicr.atr.jp/decnefpro/.

Decoder construction session

The general steps included in the decoder construction sessions are summarized in Figure 3. The fMRI task used during the decoder construction session has to be carefully designed in order to maximize the specificity and sensitivity of the target

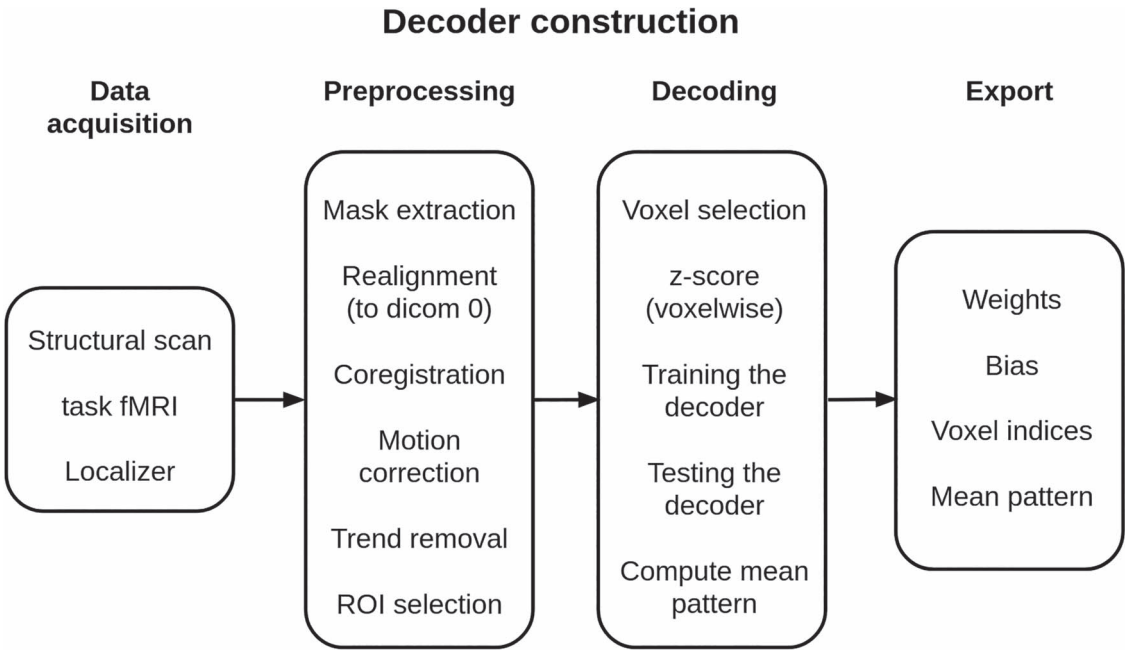


Fig. 3. General steps included in decoder construction. In order to conduct decoder construction, a structural scan and functional data from an fMRI task typically need to be acquired for each participant. A functional localizer session can also be included at this stage in order to functionally select the voxels to be used in the decoding procedure. The preprocessing steps follow standard fMRI procedure with a specific consideration for conducting steps that will also be possible to conduct online. Afterwards, the decoding steps will aim to determine the accuracy of the decoder. Once the decoder has been trained, it is important to export all the information that will be required to conduct the online decoding procedure.

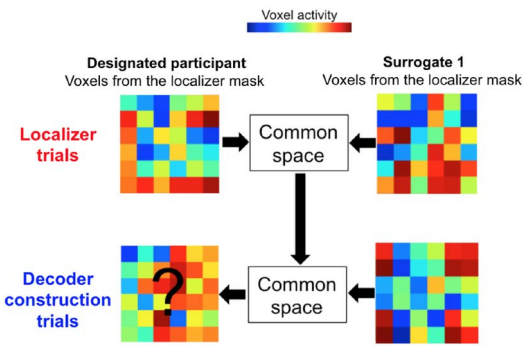
decoder. For this purpose, multiple reviews of the application of machine learning to fMRI can be consulted. Simulations can also be carried out in order to determine the optimal decoding parameters (Oblak *et al.*, 2019). Typically, two categories of stimuli (e.g. negative and positive valence) are selected in order to isolate the desired cognitive process. Here, it is important to keep in mind that the control category is as important as the target category because it will influence the specificity of the decoder. For the purpose of creating decoders of the most commonly feared animals, we previously presented participants with 3600 images of 40 different categories of animals and objects. This image database is available at https://bicr.atr.jp/decnefpro/ with psychoPy (Peirce, 2007) code that can be used to conduct the decoder construction session.

The preprocessing of the decoder construction session follows relatively standard practices in the field (see Figure 3). However, special consideration should be given to using processing steps that can be recreated in real time. This is important because the online decoding must mirror as closely as possible the preprocessing used for decoder construction. For instance, functional images can be realigned either to the first or the mean image during preprocessing as long as the same image is used to perform real-time realignment during the procedure (see below).

The target decoder is usually trained individually for each participant. This presents the advantage of conducting the decoding in the native space of each participant, which requires minimal transformation in real time. However, a decoder previously known to perform efficiently in a standard space (such as the MNI space) across participants could also be considered. Furthermore, one could also use functional alignment methods, which might be particularly useful in situations where too little training data can be obtained within-subject or when dealing with patients for whom the decoder construction session may be too aversive (see Box 1).

Box 1: functional alignment methods for decoded neurofeedback

Decoded neurofeedback relies heavily on the accuracy of the target decoder. However, some decoders can prove to be challenging to train within-subject as we can only present a limited number of trials to each participant. Furthermore, it might also be challenging for patients with specific psychological conditions to sit through the decoder construction sessions. For instance, patients presenting arachnophobia will most likely suffer great discomfort when presented with images of spiders. In these situations, modern methods of functional alignment might be useful. One such method, called hyperalignment, can be used to bring the functional data of a group of participants in alignment with the native space of a designated participant (i.e. the participant that will go through the decoded neurofeedback procedure) (see Figure) (Haxby *et al.*, 2011). Notably, this would allow for the training of decoders in the native space of a participant, using the data of a group of surrogate participants.



The first step of hyperalignment involves presenting all participants with a given task in the fMRI scanner (for instance, a localizer session). Hyperalignment operates under the assumption that this task should generate similar brain responses in all participants but that the voxels may not necessarily be aligned in space. To solve this problem, hyperalignment will perform an alignment of the functional responses to the task. This will allow for the determination of a common space in which the data of all participants can be represented (Haxby *et al.*, 2011). Importantly, this process will provide, for each participant, a transformation matrix that allows for the transformation of data from the native space to the common space.

After hyperalignment, the transformation matrix can be used to bring new data (decoder construction data in the figure) into the common space. Crucially, this transformation can also be reversed in order to bring data from the common space into the native space of a specific participant. This is an important feature of hyperalignment because it allows researchers to bring the data of surrogate participants in the native space of a designated participant through a first transformation into the common space. This process can be achieved with a large number of surrogate participants, which can greatly increase the amount of data available for training the decoders. Furthermore, this functional alignment can allow for the training of decoders to recognize brain patterns that were never expressed in the designated participants. Using this feature of hyperalignment, we previously showed that it should be possible to train decoders of feared animals without presenting participants with any aversive stimuli (Taschereau-Dumouchel *et al.*, 2018b). To avoid introducing circularity in this analysis (Kriegeskorte *et al.*, 2009), the accuracy of the decoders should be tested using data that were not included in the training of the hyperalignment or in the training of the decoder (Taschereau-Dumouchel *et al.*, 2018a)

For the purpose of decoded neurofeedback, many different implementations of hyperalignment could be explored, such as searchlight hyperalignment (Haxby *et al.*, 2011) or functional connectivity hyperalignment (Guntupalli *et al.*, 2018). Other forms of functional alignment such as the shared response model (Cohen *et al.*, 2017; Chen *et al.*, 2015) could also be considered.

The example script provided on the pyMVPA website can be used for the purpose of decoder construction with some minor adaptations (http://www.pymvpa.org/examples/hyperalignment.html). For instance, once the data are represented in the common space, it is possible to bring the data back in the native space of the designated participant using the reverse() function of the mapper as well as the transformation matrix of this specific participant.

Many algorithms can be used to conduct decoded neurofeedback. Sparse algorithms (Krishnapuram *et al.*, 2005; Yamashita *et al.*, 2008) are particularly useful for this purpose because they are more tolerant to overfitting by selecting only a few critical voxels for the classification. Some interesting resources for implementing sparse logistic regression include the SLR toolbox (https://bicr.atr.jp/~oyamashi/SLR_WEB.html), and the SMLR classifier included in the pyMVPA environment (Krishnapuram *et al.*, 2005; see http://www.pymvpa.org/generated/mvpa2.clfs.smlr.SMLR.html).

Once the decoder is trained, one important consideration is to make sure that we have everything at hand for the online decoding. This means exporting the weights determined during training as well as their associated voxel indices. In pyMVPA, the parameters of most decoders can be found by calling get_sensitivity_analyzer() (see http://www.pymvpa.org/examples/sensanas.html) while the voxel indices are stored in a feature attribute in the dataset format (see http://www.pymvpa.org/tutorial_datasets.html). It is important to recreate exactly the same prediction procedure online. As such, it is recommended to pay close attention to the code of the prediction procedure as some particularities might need to be implemented online. Also, for the purpose of monitoring the alignment of participants' brain scans in real time, it can be useful to compute the mean pattern of activity within the voxels used for decoding (see below).

The real-time decoding procedure

Figure 4 illustrates the processing steps conducted online during decoded neurofeedback. These processing steps can be achieved by distinct computer scripts executed in parallel and communicating in real time. For example, the processing and decoding steps can be independent from the visual presentation. This strategy will prevent the different components from interfering with one another and ensure a fast execution.

During the real-time decoding procedure, it is important to keep the same acquisition parameters as during decoder construction (field of view, number of slices, initial scanner alignment to the AC-PC line, etc.). Basic preprocessing first involves correcting for head movements. It is important to realign the incoming DICOM images to the functional image used as a reference during the decoder construction session. If this image is considered the 'first acquisition image' during the online procedure (sometimes referred to as the DICOM 0), then it is simple to specify the realignment of incoming images to this reference image, as registering to the first functional image is a common option in most preprocessing software. To facilitate this process, the reference image can be included in the shared folder where the DICOM images will be written during the real-time procedure. The next step involves sub-selecting only the voxels that will be used in the decoding procedure (i.e. voxel extraction). This will substantially decrease the processing time for the remaining steps as only the voxels required for decoding will be processed. This selection process can be achieved using the voxel indices extracted during decoder construction. Both of these steps can be achieved within a simple loop that continuously processes the new DICOM images as they are written to the shared directory. For this purpose, code implemented in the Matlab environment can be obtained at https://bicr.atr.jp/decnefpro/ (upon the reception of a signed agreement on terms and conditions for academic use). Functions from the most common fMRI packages can also be used as well as open-source software specifically designed for real-time processing such as OpenNFT (Koush *et al.*, 2017) or BrainIAK (Kumar *et al.*, 2019). A commercial option is also available (i.e. Turbo-BrainVoyager, Brain Innovation, the Netherlands).

In order to accommodate for the hemodynamic response function (Buckner, 1998), we have to incorporate a delay (e.g. a delay of 3 TRs when using a repetition time of 2 s). This is achieved by introducing a waiting period between the induction period and the feedback. During this period, the DICOM images are preprocessed as they are written on the hard drive, and the

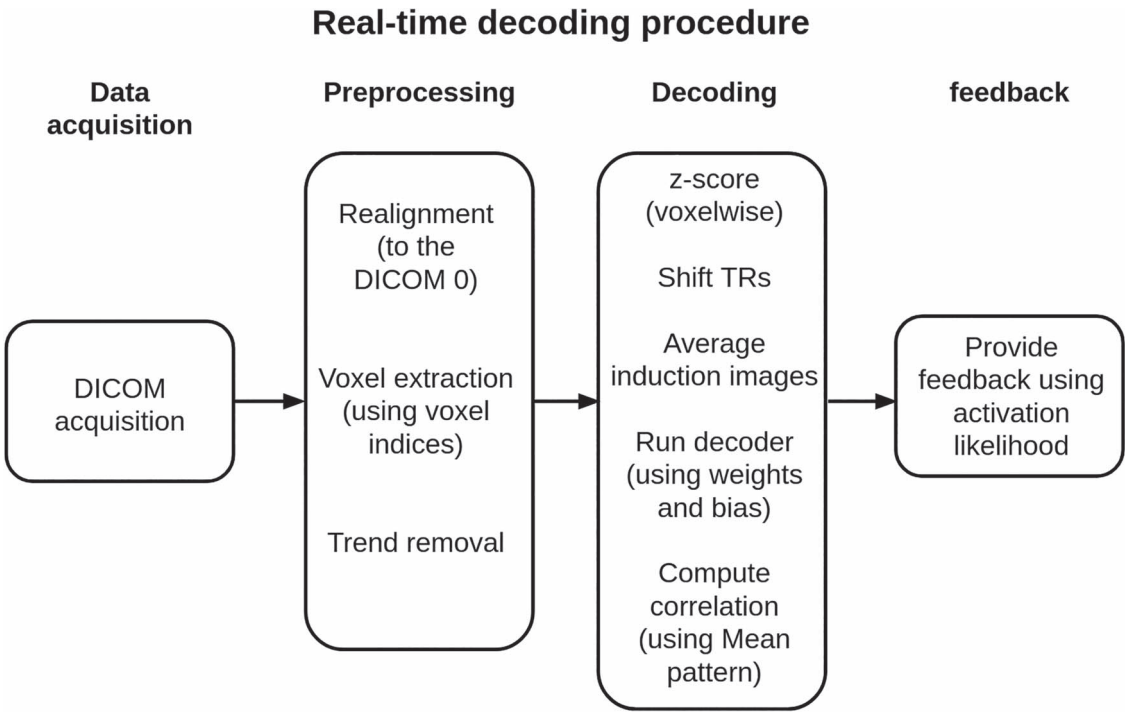


Fig. 4. General steps conducted during the online procedure of decoded neurofeedback. DICOM images are processed in real time as soon as they are available. The preprocessing steps conducted are designed to replicate as closely as possible the steps taken during decoder construction. Once all the images of the induction period are acquired and preprocessed, the real-time decoding can be achieved using the weights and bias previously determined. This step will provide us with the activation likelihood that will be displayed visually.

decoding is achieved once the last image of the waiting period is available.

The process of waiting for the DICOM images has to be flexible to accommodate possible delays in the transfer of images. As such, in the event that the DICOM images cannot be processed in time, the program should be designed to prevent providing feedback. Also, the code should be designed to perform the next steps of preprocessing and initiate decoding as soon as the last image of the waiting period is acquired.

When all the DICOM images of an induction trial are available, some additional preprocessing can be implemented. This includes the voxelwise removal of polynomial trends and the standardization of the BOLD signal. The real-time implementation of both of these processes has recently been explored (Oblak *et al.*, 2019). The results indicate that real-time processing could provide very similar results as offline processing. The best standardization appears to be achieved using all the data previously acquired during a specific run. Similarly, detrending the signal by removing the linear trends can be conducted quickly in real-time using all the images previously acquired during the run.

One important consideration for online decoding is to make sure that the realignment of the images was successful before providing the feedback. One innovative way to accomplish this is by performing, for each induction trial, a correlation between the current data and the mean activity within the voxels selected during the decoder construction (Shibata *et al.*, 2011). This information is useful, as a slight displacement of the participant will typically be associated with a clear decrease in the value of this correlation coefficient. This should be considered as a ground for repositioning the field of view of the acquisition. This can be used as complementary information to the motion parameters calculated during the realignment.

When the preprocessing steps have been completed, decoding can then be implemented. Previous studies used an induction period of 6 seconds and averaged the corresponding images in order to compute the feedback. Averaging allows for the removal of noise, thereby providing a relatively stable signal for online decoding. For decoders such as sparse logistic regression, the online decoding step is a straightforward process that involves computing the dot product of the weights and the processed signal. This information is then submitted to a logistic function in order to obtain an activation likelihood, constrained between 0 and 1, that can be communicated visually to the participant.

Decoded neurofeedback can be achieved using code implemented in the Matlab environment that can be obtained at https://bicr.atr.jp/decnefpro/ (upon the reception of a signed agreement on terms and conditions for academic use). Reinforcement scripts can also be written *de novo* using the most common software for psychological experiments such as the Psychophysics toolbox (Brainard, 1997) or psychoPy (Peirce, 2007). Some specific considerations will be needed to accommodate for potential delays in the transmission of the DICOM images (see above) and to include the processing steps during the procedure. The decoding can be achieved using the most commonly used fMRI packages as well as toolboxes such as the SLR toolbox (https://bicr.atr.jp/~oyamashi/SLR_WEB.html) and pyMVPA (Hanke *et al.*, 2009a, 2009b, 2010).

Information transmission analysis

Decoded neurofeedback experiments can be complemented by an information transmission analysis. This analysis is essentially a procedure whereby the information within a source

region is trained to predict the activation likelihood (i.e. the linearized likelihood) of the decoder. Typically, this procedure is conducted for both decoder construction and induction sessions. This way, it is possible to determine the voxels significantly associated with the activation likelihood both during decoder construction and induction. To conduct this analysis, the activation likelihood first needs to be computed and linearized for each trial (both during decoder construction and induction). Then, it will be possible to test how each source region can predict the linearized activation likelihood. An information transmission analysis relies on relatively standard analytical procedure that can be implemented in many different software and can notably be carried out in a searchlight procedure (see http://www.pymvpa.org/examples/searchlight.html) as well as in regions of interest (ROI) analyses (see Shibata *et al.*, 2011).

Further considerations

Decoded neurofeedback is still a relatively new approach, and some aspects of its implementation remain incompletely understood. For instance, many experiments are conducted in the hope of observing an increase in the activation likelihood as a result of the training. However, experiments documenting the learning effect of neurofeedback often indicate that at least four sessions of training might be required to see an increase in activation likelihood (Shibata *et al.*, 2011; Keynan *et al.*, 2019). Therefore, it is not typically expected to observe such an increase in likelihood during the first few days (Koizumi *et al.*, 2016; Cortese *et al.*, 2016a). However, this does not mean that no behavioral effects are to be expected. As discussed above, associating a reward with the (neural) occurrence of an activation pattern might be sufficient to lead to some observable changes (Koizumi *et al.*, 2016).

A related issue is that the success of the intervention depends heavily on the accuracy of the targeted decoders. This means that decoded neurofeedback can most likely be conducted only if we can build a sensitive and specific decoder of the targeted process. Unfortunately, this currently excludes some processes relevant to psychopathologies that cannot be reliably triggered in the fMRI scanner. Furthermore, while previous interventions targeted diverse brain functions (Watanabe *et al.*, 2018), the feedback was typically provided only in the visual domain. As such, further studies will be needed to determine if feedback provided in other sensory modalities (auditory, tactile, etc.) could also be used as efficiently for the purpose of decoded neurofeedback.

Another consideration pertains to the power of the intervention. Decoded neurofeedback is typically conducted in small samples because of the amount of resources necessary to implement the procedure. Therefore, this typically allows us to uncover only relatively large effects. One interesting way to improve the power of the intervention is to use within-subject designs instead of between-subject designs (Koizumi *et al.*, 2016; Cortese *et al.*, 2016b, 2017; Taschereau-Dumouchel *et al.*, 2018a). Within-subject designs are often more powerful because they intrinsically control confounding factors that otherwise have to be matched when comparing between individuals. This is also a cost-effective strategy since we do not need to use a second group of participants because of the within-subject control condition.

Future applications

Machine learning is currently a blooming field of research, and developing more sensitive and specific decoders should greatly increase the potential for decoded neurofeedback in the coming years. New technological developments such as hyperalignment (see Box 1) will also help researchers to build decoders of brain functions that are currently difficult to train. For instance, some processes such as the administration of a medication cannot be repeatedly recreated in the fMRI scanner. Using hyperalignment, it might be possible to gather the data from multiple participants and to train a multivoxel decoder of such a specific process.

Furthermore, previous studies used unidimensional feedback because it can map directly to the reward/penalty dimension used in reinforcement learning and operant conditioning. However, future experiments could consider providing multidimensional feedback to participants. For instance, one could imagine providing feedback using multidimensional scaling (Kriegeskorte *et al.*, 2008b) or using multiple brain decoders. This later possibility was notably explored in a previous experiment (Knotts *et al.*, 2019). However, in this study, activation likelihoods of multiple decoders were combined and provided to the participants through a single value. If this information was provided through multiple dimensions, it might be possible to provide more information regarding the state of the brain and to facilitate learning. Previous studies indeed indicated that multidimensional feedback can be efficiently leveraged for the purpose of reinforcement learning (Kormushev and Caldwell, 2013a,b; Niv *et al.*, 2015; Leong *et al.*, 2017) However, it is still unknown if multidimensional feedback could be successfully applied in the context of decoded neurofeedback.

Another interesting technological development is the use of parallel and cloud computing during the online procedure (Cohen *et al.*, 2017). This could potentially allow researchers to conduct complex computations in real time that are currently impossible to achieve. Ultimately, increasing the real-time computing power might even lead to an online adaptation of the target decoder. This could potentially be achieved as a function of the state of the brain in real time, which might open new possibilities for adaptations.

Yet another interesting approach is to conduct decoded EEG neurofeedback guided with fMRI data. Using this approach, it might be possible to conduct relatively inexpensive training outside of the scanner and still target very specific brain processes. Some important progress on this front has recently been made (Keynan *et al.*, 2019). However, we currently do not know if all the known properties of decoded fMRI neurofeedback could translate directly to EEG. For instance, it will be necessary to carefully determine if EEG interventions can also be conducted without any awareness of the identity of the target decoder.

Conclusion

This paper aimed to provide an overview of the methodological considerations for the implementation of decoded fMRI neurofeedback. We hope that it will help render this approach more accessible to researchers willing to adopt it. Decoded neurofeedback represents a powerful method that can be conducted in double-blind placebo-controlled settings. As such, we hope that it will become more routinely used to both study and manipulate the functioning human brain.

Funding

This manuscript was prepared with the financial support of the ImPACT Program of the Council for Science, Technology, and Innovation (Cabinet Office, Government of Japan). This work was partially supported by the contract research Grant Number JP18dm0307008 and JP17dm0107044 ('Brain machine Interface Development') under the Strategic Research Program for Brain Sciences supported by the Japan Agency for Medical Research and Development (AMED), as well as by JST ERATO under Grant Number JPMJER1801. The authors thank the US National Institute of Mental Health (Grant No. R61MH113772 to H.L.) for supporting their research leading to this work. V.T-D. is supported by a fellowship from the Fonds de Recherche du Québec—Santé (FRQS) and by a fellowship from the Canadian Institutes of Health Research (CIHR). We also thank Matthias Michel, JD Knotts and Andrew Eastwick for the helpful comments on the first version of the manuscript.

Conflict of interest

M.K. is the inventor of patents related to the neurofeedback method described here, and the original assignee of the patents is A.T.R. with which the authors are affiliated.

References

Amano, K., Shibata, K., Kawato, M., Sasaki, Y., Watanabe, T. (2016). Learning to associate orientation with color in early visual areas by associative decoded fMRI neurofeedback. *Current Biology*. **26**, P1861–1866. doi: 10.1016/j.cub.2016.05.014.

Arns, M., de Ridder, S., Strehl, U., Breteler, M., Coenen, A. (2009). Efficacy of neurofeedback treatment in ADHD: the effects on inattention, impulsivity and hyperactivity: a meta-analysis. *Clinical EEG and Neuroscience: Official Journal of the EEG and Clinical Neuroscience Society*, **40**, 180–9.

Beatty, J., Greenberg, A., Deibler, W.P., O'Hanlon, J.F. (1974). Operant control of occipital theta rhythm affects performance in a radar monitoring task. *Science*. **183**, 871–873. doi: 10.1126/science.183.4127.871.

Brainard, D.H. (1997). The psychophysics toolbox. *Spatial Vision*. **10**, 433–436 doi: 10.1163/156856897x00357.

Bray, S., Shimojo, S., O'Doherty, J.P. (2007). Direct instrumental conditioning of neural activity using functional magnetic resonance imaging-derived reward feedback. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, **27**, 7498–507.

Bu, J., Young, K.D., Hong, W., *et al.* (2019). Effect of deactivation of activity patterns related to smoking cue reactivity on nicotine addiction. *Brain*. **142**, 1827–1841. doi: 10.1093/brain/awz114.

Buckner, R.L. (1998). Event-related fMRI and the hemodynamic response. *Human Brain Mapping*. **6**, 373–377. doi: https://doi.org/10.1002/(SICI)1097-0193(1998)6:5/6%3C373::AID-HBM8%3E3.0.CO;2-P.

Chen, P.-H. (cameron), Chen, J., Yeshurun, Y., Hasson, U., Haxby, J., Ramadge, P.J. (2015). A reduced-dimension fMRI shared response model. In: Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R., editors. *Advances in Neural Information Processing Systems* **28**. Curran Associates, Inc.; pp. 460–8. Available: http://papers.nips.cc/paper/5855-a-reduced-dimension-fmri-shared-response-model.

Chiba, T., Kanazawa, T., Koizumi, A., *et al.* (2019). Current status of neurofeedback for post-traumatic stress disorder: a systematic review and the possibility of decoded neurofeedback. *Frontiers in Human Neuroscience*. **13**, 233. doi: 10.3389/fnhum.2019.00233.

Cohen, J.D., Daw, N., Engelhardt, B., *et al.* (2017). Computational approaches to fMRI analysis. *Nature Neuroscience*, **20**, 304–13.

Cortese, A., Amano, K., Koizumi, A., Kawato, M., Lau, H. (2016a). Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. *Nature Communications*. **7**, 13669. doi: 10.1038/ncomms13669.

Cortese, S., Ferrin, M., Brandeis, D., *et al.* (2016b). Neurofeedback for attention-deficit/hyperactivity disorder: meta-analysis of clinical and neuropsychological outcomes from randomized controlled trials. *Journal of the American Academy of Child and Adolescent Psychiatry*, **55**, 444–55.

Cortese, A., Amano, K., Koizumi, A., Lau, H., Kawato, M. (2017). Decoded fMRI neurofeedback can induce bidirectional confidence changes within single participants. *NeuroImage*, **149**, 323–37.

Cortese, A., De Martino, B., Kawato, M. (2019). The neural and cognitive architecture for learning from a small sample. *Current Opinion in Neurobiology*, **55**, 133–41.

Craske, M.G., Kircanski, K., Zelikowsky, M., Mystkowski, J., Chowdhury, N., Baker, A. (2008). Optimizing inhibitory learning during exposure therapy. *Behaviour Research and Therapy*, **46**, 5–27.

deBettencourt, M.T., Cohen, J.D., Lee, R.F., Norman, K.A., Turk-Browne, N.B. (2015). Closed-loop training of attention with real-time brain imaging. *Nature Neuroscience*, **18**, 470–5.

de Charms, R.C., Maeda, F., Glover, G.H., *et al.* (2005). Control over brain activation and pain learned by using real-time functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 18626–31.

Evans, J.R., Abarbanel, A. (1999). *Introduction to Quantitative EEG and Neurofeedback*. Elsevier, San Diego.

Faller, J., Cummings, J., Saproo, S., Sajda, P. (2019). *Regulation of Arousal via on-Line Neurofeedback Improves Human Performance in a Demanding Sensory-Motor Task*. doi:10.1101/428755.

Fetz, E.E. (1969). Operant conditioning of cortical unit activity. *Science*. **163**, 955–958. doi: 10.1126/science.163.3870.955.

Gharabaghi, A., Naros, G., Khademi, F., *et al.* (2014). Learned self-regulation of the lesioned brain with epidural electrocorticography. *Frontiers in Behavioral Neuroscience*, **8**, 429.

Gonçalves, S.I., de Munck, J.C., Pouwels, P.J.W., *et al.* (2006). Correlating the alpha rhythm to BOLD using simultaneous EEG/fMRI: inter-subject variability. *NeuroImage*, **30**(1), 203–13.

Guntupalli, J.S., Feilong, M., Haxby, J.V. (2018). A computational model of shared fine-scale structure in the human connectome. *PLoS Computational Biology*, **14**, e1006120.

Hammond, D.C. (2005). Neurofeedback with anxiety and affective disorders. *Child and Adolescent Psychiatric Clinics of North America*, **14**, 105–23 vii.

Hanke, M., Halchenko, Y.O., Sederberg, P.B., Hanson, S.J., Haxby, J.V., Pollmann, S. (2009a). PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics*, **7**, 37–53.

Hanke, M., Halchenko, Y.O., Sederberg, P.B., *et al.* (2009b). PyMVPA: a unifying approach to the analysis of Neuroscientific data. *Frontiers in Neuroinformatics*, **3**, 3.

Hanke, M., Halchenko, Y.O., Haxby, J.V., Pollmann, S. (2010). Statistical learning analysis in neuroscience: aiming for transparency. *Frontiers in Neuroscience*, **4**, 38.

Hardt, J.V., Kamiya, J. (1978). Anxiety change through electroencephalographic alpha feedback seen only in high anxiety subjects. *Science*, **201**, 79–81.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, **293**, 2425–30.

Haxby, J.V., Swaroop Guntupalli, J., Andrew, C.C., *et al.* (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, **72**, 404–16.

Haxby, J.V., Connolly, A.C., Swaroop Guntupalli, J. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annual Review of Neuroscience*. **37**, 435–456. doi: 10.1146/annurev-neuro-062012-170325.

Kamiya, J. (1968). Conscious control of brain waves. *PsycEXTRA Dataset*. **1**, 56–60. doi: 10.1037/e400092009-006.

Kamiya, J. (2011). The first communications about operant conditioning of the EEG. *Journal of Neurotherapy*. **15**, 65–73. doi: 10.1080/10874208.2011.545764.

Keynan, J.N., Cohen, A., Jackont, G., *et al.* (2019). Electrical fingerprint of the amygdala guides neurofeedback training for stress resilience. *Nature Human Behaviour*, **3**, 63–73.

Knotts, J.D., Cortese, A., Tascherau-Dumouchel, V., Kawato, M., Lau, H. (2019). Multivoxel patterns for perceptual confidence are associated with false color detection. doi: 10.1101/735084.

Kober, S.E., Wood, G., Kurzmann, J., *et al.* (2014). Near-infrared spectroscopy based Neurofeedback training increases specific motor imagery related cortical activation compared to sham feedback. *Biological Psychology*, **95**, 21–30.

Koizumi, A., Amano, K., Cortese, A., *et al.* (2016). Fear reduction without fear through reinforcement of neural activity that bypasses conscious exposure. *Nature Human Behaviour*. **1**, 0006. doi: 10.1038/s41562-016-0006.

Kormushev, P., Caldwell, D.G. (2013a). *Comparative Evaluation of Reinforcement Learning with Scalar Rewards and Linear Regression with Multidimensional Feedback*. September. Available: http://hdl.handle.net/10044/1/26087.

Koush, Y., Ashburner, J., Prilepin, E., *et al.* (2017). OpenNFT: an open-source python/Matlab framework for real-time fMRI neurofeedback training based on activity, connectivity and multivariate pattern analysis. *NeuroImage*, **156**, 489–503.

Kriegeskorte, N., Mur, M., Bandettini, P. (2008a). Representational similarity analysis—connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, **2**, 4.

Kriegeskorte, N., Mur, M., Douglas, A.R., *et al.* (2008b). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, **60**, 1126–41.

Kriegeskorte, N., Kyle Simmons, W., Bellgowan, P.S.F., Baker, C.I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, **12**, 535–40.

Krishnapuram, B., Carin, L., Figueiredo, M.A.T., Hartemink, A.J. (2005). Sparse multinomial logistic regression: fast algorithms and generalization bounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**, 957–68.

Kumar, M., Ellis, C.T., Lu, Q., *et al.* (2019). BrainIAK tutorials: user-friendly learning materials for advanced fMRI analysis. doi: 10.31219/osf.io/j4sbc.

Leong, Y.C., Radulescu, A., Daniel, R., DeWoskin, V., Niv, Y. (2017). Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron*, **93**, 451–63.

Mennella, R., Patron, E., Palomba, D. (2017). Frontal alpha asymmetry neurofeedback for the reduction of negative affect and anxiety. *Behaviour Research and Therapy*, **92**, 32–40.

Niv, Y., Daniel, R., Geana, A., *et al.* (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, **35**, 8145–57.

Oblak, E.F., Lewis-Peacock, J.A., Sulzer, J.S. (2017). Self-regulation strategy, feedback timing and hemodynamic properties modulate learning in a simulated fMRI neurofeedback environment. *PLoS Computational Biology*, **13**, e1005681.

Oblak, E.F., Sulzer, J.S., Lewis-Peacock, J.A. (2019). A simulation-based approach to improve decoded neurofeedback performance. *NeuroImage*, **195**, 300–10.

Oblak, E. Lewis-Peacock, J., Sulzer, J. 2020, Differential neural plasticity of individual fingers revealed by fMRI neurofeedback. *Biorxiv*, https://www.biorxiv.org/content/10.1101/2020.03.02.973586v2.abstract

Peirce, J.W. (2007). PsychoPy—psychophysics software in python. *Journal of Neuroscience Methods*. **162**, 8–13. doi: 10.1016/j.jneumeth.2006.11.017.

Ramot, M., Gonzalez-Castillo, J. (2019). A framework for offline evaluation and optimization of real-time algorithms for use in neurofeedback, demonstrated on an instantaneous proxy for correlations. *NeuroImage*, **188**, 322–34.

Ramot, M., Kimmich, S., Gonzalez-Castillo, J., *et al.* (2017). Direct modulation of aberrant brain network connectivity through real-time NeuroFeedback. *eLife*, **6**, e28974. doi: 10.7554/eLife.28974.

Schabus, M., Griessenberger, H., Gnjezda, M.-T., Heib, D.P.J., Wislowska, M., Hoedlmoser, K. (2017). Better than sham? A double-blind placebo-controlled neurofeedback study in primary insomnia. *Brain*. **140**, 1041–1052. doi: 10.1093/brain/awx011.

Schönenberg, M., Wiedemann, E., Schneidt, A., *et al.* (2017). Neurofeedback, sham neurofeedback, and cognitive-behavioural group therapy in adults with attention-deficit hyperactivity disorder: A triple-blind, randomised, controlled trial. *The Lancet. Psychiatry*, **4**, 673–84.

Sepulveda, P., Sitaram, R., Rana, M., Montalba, C., Tejos, C., Ruiz, S. (2016). How feedback, motor imagery, and reward influence brain self-regulation using real-time fMRI. *Human Brain Mapping*, **37**, 3153–71.

Shibata, K., Watanabe, T., Sasaki, Y., Kawato, M. (2011). Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation. *Science*. **311**, 1413–1415. doi: 10.1126/science.1212003.

Shibata, K., Watanabe, T., Kawato, M., Sasaki, Y. (2016). Differential activation patterns in the same brain region led to opposite emotional states. *PLoS Biology*, **14**, e1002546.

Shibata, K., Lisi, G., Cortese, A., Watanabe, T., Sasaki, Y., Kawato, M. (2019). Toward a comprehensive understanding of the neural mechanisms of decoded neurofeedback. *NeuroImage*, **188**, 539–56.

Strohminger, N., Gray, K., Chituc, V., Heffner, J., Schein, C., Heagins, T.B. (2016). The MR2: a multi-racial, mega-resolution database of facial stimuli. *Behavior Research Methods*, **48**, 1197–204.

Taschereau-Dumouchel, V., Cortese, A., Toshinori, C., Knotts, J.D., Kawato, M., Lau, H. (2018a). Towards an unconscious neural reinforcement intervention for common fears. *Proceedings of the National Academy of Sciences of the United States of America*, **115**, 3470–5.

Taschereau-Dumouchel, V., Liu, K.-Y., Lau, H. (2018b). Unconscious psychological treatments for physiological survival circuits. *Current Opinion in Behavioral Sciences*, **24**, 62–8.

Taschereau-Dumouchel, V., Kawato, M., Lau, H. (2019). Multivoxel pattern analysis reveals dissociations between subjective fear and its physiological correlates. *Molecular Psychiatry*, October. doi: 10.1038/s41380-019-0520-3.

Taschereau-Dumouchel, V., Chiba, T., Koizumi, A., Kawato, M., Lau, H. (2020). Multivoxel neural reinforcement changes resting-state functional connectivity within the threat regulation network. *bioRxiv*. doi: 10.1101/2020.04.03.021956.

Thibault, R.T., Lifshitz, M., Birbaumer, N., Raz, A. (2015). Neurofeedback, self-regulation, and brain imaging: clinical science and fad in the service of mental disorders. *Psychotherapy and Psychosomatics*, **84**, 193–207.

Van Doren, J., Arns, M., Heinrich, H., Vollebregt, M.A., Strehl, U., Loo, S.K. (2019). Sustained effects of neurofeedback in ADHD: a systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, **28**, 293–305.

Watanabe, T., Sasaki, Y., Shibata, K., Kawato, M. (2017). Advances in fMRI real-time neurofeedback. *Trends in Cognitive Sciences*, **21**, 997–1010.

Watanabe, T., Sasaki, Y., Shibata, K., Kawato, M. (2018). Advances in fMRI real-time neurofeedback: (Trends in Cognitive Sciences 21, 997-1010, 2017). *Trends in Cognitive Sciences*, **22**, 738.

Weiskopf, N. (2012). Real-time fMRI and its application to neurofeedback. *NeuroImage*. **62**, 682–692. doi: 10.1016/j.neuroimage.2011.10.009.

Yamashita, O., Sato, M.-A., Yoshioka, T., Tong, F., Kamitani, Y. (2008). Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns. *NeuroImage*, **42**, 1414–29.

Young, K.D., Siegle, G.J., Zotev, V., *et al.* (2017). Randomized clinical trial of real-time fMRI amygdala neurofeedback for major depressive disorder: effects on symptoms and autobiographical memory recall. *The American Journal of Psychiatry*, **174**, 748–55.