1 Intellectual Merit

Social interactions are critical for survival, and lie at the foundations of cooperation and cumulative culture. The human ability to interact with others relies on social perception: our capacity to recognize people's identities, emotions, and actions [7, 23]. Previous work has investigated neural representations of changeable properties of people (i.e. expressions, actions) and more lasting properties (i.e. identity). A traditional view proposes that recognition of face identity is performed by a ventral temporal stream, and recognition of expressions by a lateral stream [10, 33]. However, recent findings challenge this view: identity can be decoded from brain regions in the lateral stream [3, 16], and expressions from regions in the ventral stream [61]. In contrast with the traditional view, we hypothesize that recognition of face identity and facial expressions are performed by shared neural mechanisms, and that this is due to computational efficiency: representations of identity could aid expression recognition by identifying aspects of a face image that are not due to identity.

Body-selective regions are adjacent to face-selective regions, and are also organized into two streams. We hypothesize that like face identity and expressions, body identity and actions are represented by common neural mechanisms. Finally, we hypothesize that rather than differing in terms of the content they represent (identity vs expressions or actions), ventral and lateral streams differ in terms of the type of information they process: the former might process static shape features, while the latter might process optic flow.

The project we propose aims to test these hypotheses using behavioral studies, functional magnetic resonance imaging (fMRI), deep networks, and multivariate connectivity. This project is central for my career: it tests hypotheses stemming from my previous work [3, 51], it applies new computational techniques I have developed ([4, 5, 45]), and it establishes the foundations to study the interplay between social perception and the acquisition of person knowledge - linking the past, present and future of my research program.

Objective 1: to study the computational and neural architecture of face perception. We will study whether training deep networks to recognize expressions leads to the spontaneous formation of representations of identity (we will refer to this phenomenon as 'complementarity'). This would suggest that ventral and lateral streams in the human brain might not be specific respectively for identity and expressions. Instead, they might be specialized for static features and optic flow. We will use facial expression videos and deep analysis-by-synthesis [68] to train a two-stream deep network [60] to recognize expressions. One of the streams will process individual frames ('static stream'), and the other optic flow ('dynamic stream') [60]. First, we will test whether, as successive layers of the network yield increasingly accurate expression recognition, labeling of identity also improves, despite the network was never explicitly trained to recognize identity. Second, we will use fMRI and encoding models [50] to test whether the deep net features accurately predict fMRI data, and whether features in the deep net's static stream better predict fMRI responses in the ventral stream, and features in the deep net's dynamic stream better predict fMRI responses in the lateral stream.

Objective 2: to study the computational and neural architecture of body and action perception. Is complementarity a more general principle of social perception? We will study its extent and limits investigating whether algorithms trained to recognize actions develop increasingly accurate representations of a body's identity. Perception of the body in motion can be used to recognize a person's identity [52]. We will train a two-stream network [60] with the Kinetics-700 database, and we will test whether in both the static and dynamic streams, as representation yield increasingly accurate labeling of actions, labeling of identity also improves. We will then test with encoding models [50, 49] whether features in the two streams in the deep network differentially predict responses in the ventral and lateral streams of the human brain, using fMRI data from 30 healthy human adults and a publicly available fMRI dataset [31].

Objective 3: to characterize the multivariate interactions between social perception regions. While Objectives 1 and 2 focus on representational content, in Objective 3 we will study how regions in the two streams interact. We will use a new analysis technique - Multivariate Pattern Dependence (MVPD, [4]) - which describes brain regions' responses as trajectories in a multivariate representational space, and learns a mapping between them. The most recent variant of MVPD (based on artificial neural networks) will be applied to the analysis of fMRI responses measured while participants observe 1) controlled stimuli, and 2) whole movies. In addition, we will use a new MVPD-based approach to identify brain regions that integrate information from both the ventral and lateral temporal streams.

I expect this project will lead to a **transformative, computational understanding of the mechanisms for social perception** and of how they are implemented in the brain, serving as a stepping stone to study social cognition more broadly, and providing a new foundation to gain insights into social perception

deficits.

2 Research plan

2.1 Background Information

According to a traditional account [10, 33], the neural mechanisms for face perception are organized into two neural pathways: a ventral temporal pathway specific for identity (including the occipital face area, OFA [29]; and the fusiform face area, FFA [37]), and a lateral temporal pathway specific for expressions (including the posterior superior temporal sulcus, pSTS [48]). However, several recent findings challenge this account. Information about face identity has been decoded from pSTS [3, 16, 32], and information about the valence of expressions has been decoded from ventral temporal regions previously implicated in the recognition of identity [61, 40]. Importantly, a case study of a patient with pSTS damage revealed a deficit affecting not only the recognition of expressions, but also the recognition of face identity with robustness across expression changes [20]. In this project, we will investigate whether joint representations of identity and expressions arise naturally in artificial neural networks, and use the resulting computational models to predict neural responses in humans measured with fMRI.

Like expressions and face identity jointly contribute to the appearance of dynamic faces, actions and body identity jointly contribute to the appearance of behaving people. Different individuals have limbs of different lengths, which rotate around the joints tracing identity-specific trajectories. Person identity can be recognized from the moving body even when face information is obscured [52]. Like the neural mechanisms for face processing, the mechanisms for body processing are organized into a ventral stream, including the extrastriate body area (EBA, [17]) and the fusiform body area (FBA, [59]), and a lateral stream, including the body-selective pSTS [63, 12]. Actions can be decoded in body-selective regions [30]. Furthermore, a region involved in action recognition, the lateral occipital temporal complex (LOTC, [46, 66, 14]), likely overlaps with body-selective regions. Inspired by these findings, we will study whether representations of body identity also arise spontaneously in deep networks trained to recognize actions, and we will use features from the deep networks to predict fMRI responses, determining whether representing orthogonal properties within common brain regions is a more general principle of organization of social perception that shapes both face and body recognition.

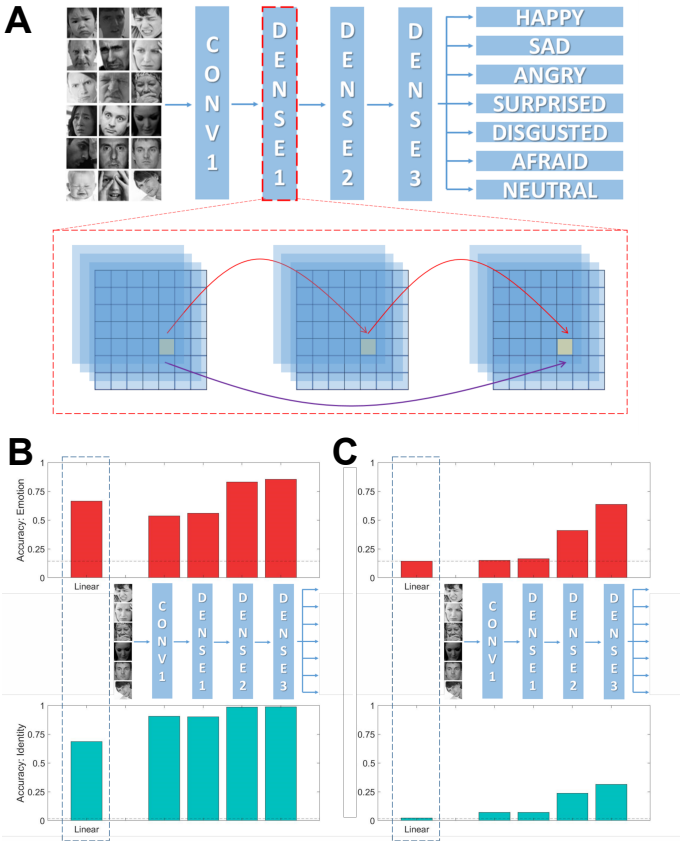

Figure 1: Performance of different layers of a deep neural network trained to label expressions at labeling expressions (red bars) and identity (green bars) in a novel dataset. A) Structure of the network; B) Accuracy when all face viewpoints are included in the training of the linear layer; C) Accuracy when training and testing the linear layer across different viewpoints.

Discovering joint representations of face identity and expressions, and of body and actions, would leave open the question of the distinct roles of the ventral and lateral streams. We hypothesize that the lateral stream represents dynamic features of the stimuli, such as optic flow, while the ventral stream represents static features, such as shape. This hypothesis is fundamentally different from the proposal of separate streams for identity and expressions: several studies indicate that dynamic features contribute to the recognition not only of changeable properties like expressions and actions, but also of more lasting properties like identity ([52, 15],

see [69] for a review).

Neuroimgaging studies found stronger responses to dynamic than static face stimuli in right pSTS [21, 56]. Furthermore, pSTS receives projections from the middle temporal visual area (MT) [24], an area representing motion information. These observations suggest a role for pSTS in processing dynamic properties, but leave open the questions of what properties pSTS represents, and how they are computed from videos. Recent advances in deep networks [41, 67] offer a computational framework to investigate these questions. Recent deep network architectures for video recognition employ two processing streams: one processing individual frames ('static' stream), the other processing optic flow ('dynamic' stream) [60, 13]. We will take advantage of two-stream deep networks and encoding models [50] to investigate 1) whether joint representations of expression and identity and of the body and actions emerge in both steams of deep network models, and 2) whether the dynamic stream of the networks better predicts responses in the lateral temporal stream in the brain.

2.2 Preliminary Results

2.2.1 Complementary Representations of Identity and Expressions in One-stream Networks

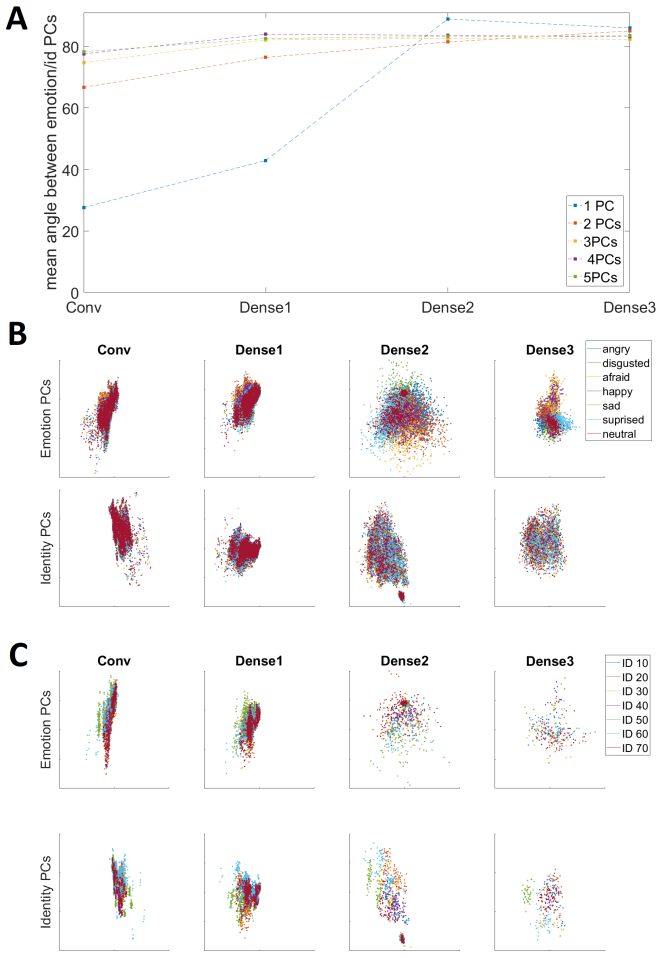

Figure 2: A) Principal components (PCs) in feature space explaining variation in expression and identity become increasingly orthogonal from layer to layer; B) Emotions are clustered in the space spanned by PCs explaining variation in emotion, but not in the space spanned by PCs expaining variation in identity; C)Identiies are clustered in the space spanned by PCs explaining variation in identity, but not in the space spanned by PCs expaining variation in emotions.

Understanding what aspects of a face image are due to a person's identity could enable an observer to avoid errors due to attributing those aspects to the person's expression. If this is the case, training a deep network to label expressions could lead to the spontaneous emergence of representations that support identity recognition. Such 'complementarity' of identity and expressions could offer a computational explanation for the observation that both identity and expressions can be decoded from the same brain regions [3, 61].

In preliminary research, we tested the spontaneous emergence of identity representations in deep networks processing static images. We trained a densely connected convolutional network (DenseNet, [34]) to label facial expressions using the fer2013 dataset. The network learned to label expressions with human-level accuracy (63%, [28]). We then tested the network's ability to label expressions and identity in a novel dataset, the Karolinska Directed Emotional Faces (KDEF).

First, we tested that the network could generalize to the new dataset, calculating the accuracy at labeling expressions in KDEF images. To evaluate the performance of individual layers, we 'froze' the network's weights (that is, we prevented them from learning further), and used a subset of the KDEF images to train a new linear layer that takes as inputs the features in one of the frozen network's layers, and generates as output an expression label (surprised, angry, sad, disgusted, scared, happy, or neutral). We used the remaining images in the KDEF dataset to compute the network's accuracy at labeling expressions. We repeated this procedure for each layer in the network, yielding the plot shown in Figure 1 (red bars). The

procedure was performed twice: once training the linear layer with a subset of the identities, and testing it with novel identities (Figure 1, left), the second time training the linear layer with a subset of the viewpoints, and testing on a new viewpoint (Figure 1, right). Importantly, the linear layer trained with a subset of KDEF stimuli cannot learn new nonlinear features, and thus needs to rely on the nonlinear features it receives as input from the frozen DenseNet layer trained with the fer2013 dataset.

As expected, accuracy at labeling expressions increased when using the features from later layers in the frozen network, reaching the highest accuracy at the last layer (Dense 3, see Figure 1). Critically for our hypothesis, we also expected that accuracy for identity labeling would correspondingly improve from early to later layers in the network. In line with the prediction, the accuracy for identity labeling increased when using features from later layers in the network, in correspondence with the increase in accuracy for expressions (Figure 1). Furthermore, identity labeling using features from the network trained to label expressions outperformed the accuracy obtained training a linear layer directly on the images (baseline).

In addition, dimensions in feature space encoding identity and expressions became increasingly orthogonal from layer to layer (Figure 2 A). This finding suggests that increasing accuracy for identity labeling does not emerge in the network trained to label emotions because identity and emotion share common features, but rather, that features encoding identity and expression become increasingly orthogonal in the network. In line with this observation, in a lower-dimensional space spanned by principal components capturing variation across emotions, stimuli were found to increasingly cluster by emotion from layer to layer, but not by identity (Figure 2 B). Viceversa, in a lower-dimensional space spanned by principal components capturing variation across identities, stimuli were found to increasingly cluster by identity from layer to layer, but not by emotion (Figure 2 C) In the first Objective, we will extend these results investigating complementarity in two-stream networks, and we will study whether the two streams of the networks differentially predict responses in the ventral and lateral streams in the human brain. In the second Objective, we will test whether complementarity is a more general principle in social perception, investigating the recognition of the body and of the actions it performs.

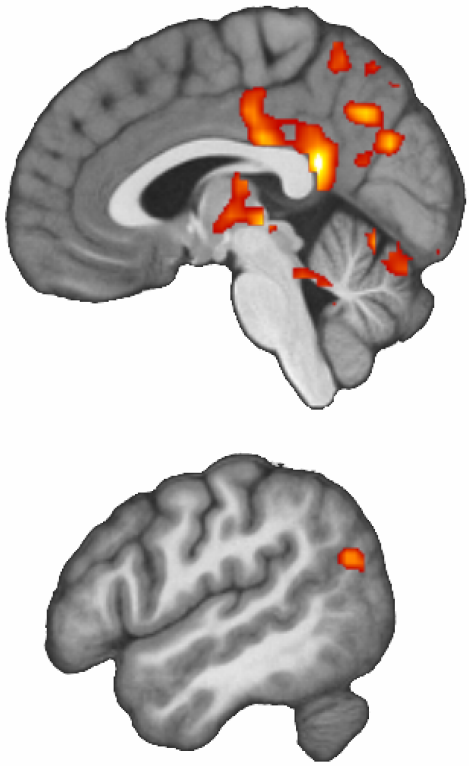

Figure 3: Brain regions showing evidence of integration of information across multiple category-selective networks. Heat maps denote the difference between the proportion of variance explained in each voxel using as predictors response patterns in all category selective networks, minus the proportion of variance explained using as predictors only regions selective for the one category among the four categories tested (faces, bodies, objects, scenes) yielding the best prediction for that voxel.

2.2.2 Multivariate Integration of Information Across Multiple Streams

In the third Objective, we will use multivariate pattern dependence (MVPD, [4]) to study the interactions between brain regions engaged in social perception. MVPD is a new method to study the interactions between brain regions in terms of their multivariate patterns of response (see also [5, 45]). We will identify brain regions that integrate information from both the ventral and lateral streams, investigating whether their responses are better predicted by MVPD using as input the combination of the response patterns in both streams than by MVPD using as input the single stream that provides the most accurate prediction.

We have tested this approach to investigate a problem with similar formal structure: identifying brain regions that integrate information across brain regions selective for different categories. Using the StudyForrest dataset [31], we defined regions of interst (ROIs) for face-selective, body-selective, artifact-selective, and scene-selective regions. We then searched for brain regions whose responses are better predicted by regions

selective for these 4 categories combined, than by the single category yielding the most accurate predictions in isolation. This analysis identified the posterior and middle cingulate gyrus, the thalamus, and the angular gyrus ($p < 0.05$ voxelwise corrected, assessed with Statistical Nonparametric Mapping, Figure 3).

2.3 Objective 1: Computational and Neural Architecture of Face Perception

2.3.1 Complementary representations of identity and expressions in dynamic stimuli

We hypothesize that complementarity between identity and expressions is not only present for representations of static face images, but also for representations of dynamic faces (Hypothesis 1.1). To test this hypothesis, we will train two-stream neural networks [60] to label expressions in dynamic videos, and we will test whether the networks learn features that contribute to identity recognition, not only in the static but also in the dynamic stream.

Generating realistic, controlled face dynamics with deep analysis-by-synthesis. Databases of dynamic face stimuli do not approach the size of databases of static face images. The extended Cohn-Kanade database (CK+) [47] contains 327 labeled videos, the Oulu-CASIA database [71] contains 480 videos, and the MMI database contains 2900 videos [53], as compared to the fer2013 dataset which contains over 30,000 images. We propose to mitigate this issue generating a large number of expression videos with realistic dynamics thanks to a novel 2-stage strategy based on deep analysis-by-synthesis [68].

In the first stage, a rendering software (FACSHuman, [25], see Figure 4) will be used to generate 30,000 images of different identities displaying different expressions. For each image, parameters for the identity and for the activation of Action Units (AUs) will be extracted from a uniform distribution. The images obtained will be given as inputs to a deep network, which will be trained to produce as output the parameters that generated the image. In the second stage, labeled naturalistic videos of facial expressions from the MMI database [53] will be decomposed into frames, and given as input to the deep network that was trained with the generative model. For each frame, the network will output values for the activation of each of the AUs, yielding a timecourse of naturalistic AU dynamics. We will then use FACSHuman to apply the extracted dynamics to a variety of artificial identities, producing a large database of expression videos. The database and the trained network will be made publicly available to the research community on Github.

A known challenge for research on facial expressions is that most stimuli consist of acted expressions, which might differ from spontaneous expressions 'in the wild'. To mitigate this issue, for the emotions happiness, surprise, and disgust, we will rely on the spontaneous expressions available in the MMI database. For anger, fear, and sadness, these emotions were not induced in the MMI database due to ethical concerns, therefore we will rely on the acted videos.

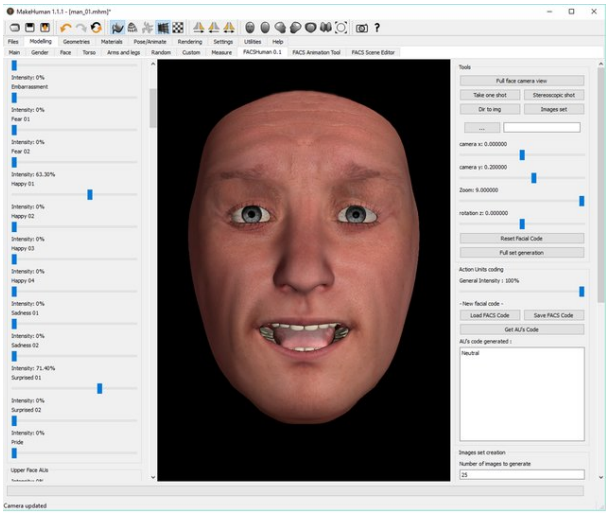

Figure 4: Facshuman software interface.

If the videos produced with analysis-by-synthesis were not sufficient to expand the MMI database and train the two-stream network to achieve high accuracy, we will complement them with the extended Cohn-Kanade database (CK+) [47] and the Oulu-CASIA database [71]. If that were also not sufficient, we will generate additional videos using generative adversarial networks (GAN, [27]). Specifically, after extracting AU timecourses from the MMI database, we will use a generator network to produce candidate AU timecourses, and a discriminator network to distinguish between AU timecourses extracted from MMI videos and timecourses produced by the generator. The generator and the discriminator will be trained in an adversarial fashion, until the generator can produce novel AU timecourses that the discriminator can no longer distinguish from real AU timecourses extracted from MMI.

Testing the complementarity of identity and expression for static and dynamic features. A two-stream convolutional network [60] will be trained to label facial expressions using the MMI database augmented with the large database of realistic dynamic expressions generated with deep analysis-by-synthesis.

We will then use the approach described in detail in the Preliminary Results to test whether from layer to layer, as features yield increasing accuracy for labeling expressions, they also yield increasing accuracy for labeling identity. This procedure will be performed separately for the two streams of the network, to test whether increasingly accurate identity labeling is observed also in the dynamic stream. As in [51], we will use principal component analysis (PCA) to test whether labeling of expressions and labeling of identity rely on features that become increasingly orthogonal from layer to layer. This analysis will enable us to determine whether complementarity of identity and expression extends to dynamic representations. In future research, beyond the scope of this proposal, we will test the symmetrical case in which the network is trained to label identity, and tested to label expressions.

If the accuracy of two-stream networks were not sufficiently high when using a simple feedforward architecture, we will use densely connected networks (DenseNets, [34]), so that each of the streams in the two-stream network is densely connected. DenseNets have connections between non-adjacent layers, and have been shown to achieve high accuracy in a variety of tasks [34]. We have used DenseNets in previous work with good results [51].

2.3.2 Research direction: neural representations of static and dynamic face features

The presence of information about both identity and expressions in both the FFA [6, 2, 61] and pSTS [3, 32, 54, 61] indicates that the ventral and lateral streams for face processing are not distinguished by whether they represent identity or expressions. Instead, the ventral stream might encode static properties of faces, and the lateral stream might encode dynamic properties. We will test this hypothesis and evaluate the two-stream deep network model we trained in the previous stage of the project as an account of how these properties are computed.

Considering the the existing empirical evidence in light of the two-stream deep network model, we hypothesize that the static stream of the model will predict fMRI responses in OFA and FFA better than the dynamic stream (Hypothesis 1.2); and the dynamic stream will predict fMRI responses in pSTS better than the static stream (Hypothesis 1.3). Recent research [57] shows that TMS to OFA reduces pSTS responses to static faces but not to dynamic faces, suggesting that, in addition to dynamic face information, pSTS receives as input static face information from OFA. For this reason, we hypothesize that features from the static stream of the deep network will predict additional variance in pSTS that is not explained by the features in the dynamic stream alone (Hypothesis 1.4). We will test these hypotheses using encoding models [50] to predict response patterns in pSTS, OFA and FFA during the perception of controlled face videos (in Experiments 1.1-1.2). To this end, Experiment 1.2 will play a critical role, enabling us to obtain the needed statistical power to train the encoding models thanks to a stimulus set consisting entirely of facial expressions. We will then test the models' ability to predict responses to complex videos where dynamic faces appear in context (in Experiment 1.3).

Experiment 1.1: functional localizer. A total of 30 adults (age 18-50, approximately half female) will be recruited to participate to Experiments 1.1 and 1.2. Experimental procedures will be approved by Boston College's Institutional Review Board, and each participant will provide informed consent before taking part to the experiment. Experiment 1.1 will consist of two runs. In the first run, participants will be shown 6 types of stimuli: static images of faces, static images of bodies, videos of faces, bodies (actions), artifacts and scenes. For each stimulus type, 4 blocks of 20 seconds duration will be shown, separated by 6 seconds of fixation, leading to a total duration of approximately 11 minutes. In the second run, participants will be shown point-light displays [8] of walking humans, point-lights moving in random directions, and static point-lights. For each of the 3 stimulus types, 5 blocks of 20 seconds duration will be shown, separated by 6 seconds of fixation, leading to a total duration of approximately 7 minutes. In both runs, participants will perform a 1-back task, and in 10% of the trials two identical stimuli will be shown in a row.

Experiment 1.2: controlled stimuli. In experiment 1.2, we will test hypotheses 1.2-1.4 using videos of facial expressions in controlled settings from the MMI database. The experiment will consist of 6 runs, a 9 minutes resting state scan, and a 5 minutes anatomical scan (MPRAGE). During each of the 6 runs, participants will be shown 2 second long videos of facial expressions: 10 for each emotion label (anger, sadness, fear, disgust, surprise, happiness, neutral). Videos will be presented in randomized order. Participants will be asked to press a button whenever a neutral video is shown. 5 face identities will be shown, each contributing two videos for each emotion. Each video will be followed by a jittered intertrial interval of 4-8 seconds extracted from a uniform distribution, leading to a total run duration of approximately 9 minutes.

Experiment 1.3: dynamic expressions in context. In experiment 1.3 we will test the three hypotheses (1.2-1.4) using responses to videos of facial expressions in context, where a face can appear concurrently with other faces and objects, and multiple facial expressions can be visible simultaneously on the screen. For this purpose, we will use the publicly available StudyForrest dataset, and specifically the data from the 15 subjects who saw the visual 'Forrest Gump' movie [31]. The StudyForrest dataset includes a localizer for category-selective regions, which will be used to identify faces-selective regions.

FMRI data analysis. To maximize the reproducibility of the results, all data will be preprocessed with fmriprep [18] using a docker container converted to a singularity image to run on Boston College's Sirius cluster, following a pipeline that has been previously used in my laboratory [45]. Based on previous tests [45] denoising will be performed with aCompCor. Additionally, timeponts with motion and global signal outliers will be removed from the data. Functional localizers will be analyzed with a General Linear Model implemented in FSL FEAT [65], with boxcar predictors for different stimulus types, convolved with a standard haemodynamic response function (HRF).

For the analysis of responses to facial expressions, the videos will be given as inputs to the two-stream network trained in the previous stage of Objective 1; the activations of hidden units will be convolved with a standard HRF and used as predictors in an encoding model [50]. The encoding model will be trained and tested on independent data using a leave-one-run-out cross-validation. To prevent overfitting, if needed, we will use regularization (i.e. elastic net [72]). The proportion of variance explained in independent data will be used as a metric to evaluate different models. For each ROI (OFA, FFA, pSTS) we will compare three models: one using as predictors the features extracted from the static stream of the two-stream network, one using the features extracted from the dynamic stream, and one using both. In addition to these analyses, if time allows, we will perform a whole brain analysis identifying voxels where responses are well predicted by each of the two streams of the model.

2.4 Objective 2: Computational and Neural Architecture of Body and Action Perception

2.4.1 Complementarity of Actions and Body Identity in Two-Stream Networks

Like facial expressions and face identity jointly contribute to the appearance of face videos, actions and body identity jointly contribute to the appearance of videos of behaving humans. We hypothesize that as in the case of faces, representations of actions and body identity are complementary (Hypothesis 2.1). We will test this hypothesis training a two-stream network using the Kinetics-700 Human Action Video dataset [13]: a recent database for action recognition that comprises 700 action classes, and 600 unique videos for each class ([39], see Figure 5 for example stills). We will then assess the network's ability to label actions and identities using the i3dPost dataset [26], which includes a variety of actions performed by 8 individuals filmed from different viewpoints. Critically, the i3dPost dataset includes frame-by-frame mesh reconstructions of the actions, which can be used to generate videos that discard information about the luminance and texture of clothing and force the network to rely more on motion patterns. Adopting a procedure analogous to that used for faces, we will test whether accuracy for the classification of identity increases from layer to layer, despite the nonlinear component of the network was only trained to label actions. Accuracy will also be compared to a 'baseline' linear network, and PCA will be used (as in [51]) to test whether labeling of actions and labeling of identity rely on features that become increasingly orthogonal from layer to layer.

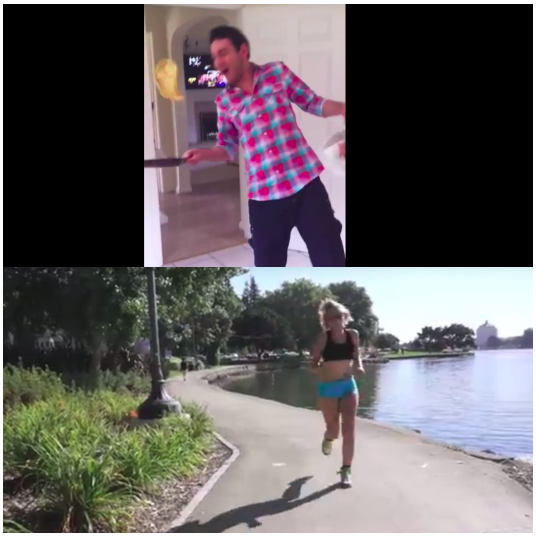

Figure 5: Example stills from the Kinetics-700 dataset, labeled as 'flipping pancake' (top) and 'jogging' (bottom).

2.4.2 Neural Representations of Static and Dynamic Action Features

We hypothesize that responses in ventral body-selective brain regions are better explained by the static stream of the deep network than by the dynamic stream (Hypothesis 2.2). We additionally predict that

the body-selective pSTS is better explained by the dynamic stream than by the static stream (Hypothesis 2.3). Finally, by analogy with expression recognition, we expect that the static stream of the deep network explains additional variance in the lateral temporal stream in the brain, that is not already explained by the dynamic stream of the deep network (Hypothesis 2.4). In Experiment 2.1 and 2.2, we will test these hypotheses using fMRI data recorded while participants watch a highly heterogeneous subset of the Kinetics-700 action videos, obtaining exceptional statistical power thanks to a varied stimulus set consisting entirely of actions. In Experiment 2.3, we will replicate the results with the publicly available StudyForrest dataset.

Experiment 2.1: Functional Localizer. Experiment 2.1 will be identical to Experiment 1.1, and will be performed in a new group of 30 healthy adults that will also complete the fMRI portion of Experiment 2.2.

Experiment 2.2: Kinetics-700. We will select 60 actions from the 700 action labels of Kinetics-700 for use in the fMRI experiment using a multi-stage procedure. First, we will choose the 200 more frequent labels using the English Corpus of Google Ngram. Subsequently, we will ask 20 participants to rate each action label from 1 to 10 based on how frequently they observe it throughout a year. These ratings will be averaged and used to select 100 frequently encountered actions. Another group of 200 participants will rate the similarity between pairs of actions. Each participant will be given an action as reference, and will be asked how similar it is to each of the other 99 (presented in randomized order). Using these data, we will generate a matrix reflecting the similarity between each pair of actions, and select a subset of 60 actions that are most distinct from each other (to include a variety of actions in the fMRI experiment). For each action label, we will select 3 videos with that label from the Kinetics-700 dataset. Thanks to this stimulus selection procedure, Experiment 2.2 will enable us to test our hypotheses using fMRI responses to a rich set of videos spanning a wide variety of actions.

A total of 30 healthy adults will take part to the fMRI portion of Experiment 2.2. The fMRI session will consists of 6 runs. In each run, 30 action videos will be shown, each lasting 10 seconds and followed by a 8 seconds intertrial interval (total run duration: 9 minutes). On 10% of the trials, a red or blue fixation cross will be shown during the intertrial interval. If the cross is red, the participant will have to press a button with the index finger if the gender of the agent is female, and with the middle finger if it is male. If the cross is blue, the participant will have to press a button with the index finger if the agent grasped an object with her/his hands, and with the middle finger if s/he didn't. Thanks to this task, participants will need to attend to both the agent and the action s/he is performing. Runs 1 and 2 will contain one video from each of the 60 action labels, as will runs 3 and 4, and runs 5 and 6, so that these pairs of runs can be used as separate folds for training and testing. All other aspects of the ordering of the videos will be randomized.

Experiment 2.3: StudyForrest. In Experiment 2.3, we will replicate the results obtained in Experiment 2.2 using the data in the publicly available StudyForrest dataset [31]. An independent group of 10 participants will be recruited to annotate the movie for the presence of actions.

FMRI data analysis. Preprocessing and denoising for experiments 2.1 and 2.2 will be performed as for experiments 1.1 and 1.2. In experiment 2.1, the functional localizer will be used to identify regions selective for bodies (EBA, FBA, body-selective pSTS), and regions responding to actions. In experiment 2.3, the category localizer will be used to identify regions selective for bodies. The movie data will be used to identify regions responding to actions: we will generate a predictor that has value of 1 for timepoints when more than 50% of the 10 raters reported an action, and 0 otherwise. Action ROIs will be identified with a contrast of this predictor vs baseline (the other parts of the video). For each ROI, BOLD responses will be predicted with a regularized encoding model [50] using as predictors either the features from the static stream of the two-stream networks, the features from the dynamic stream of the two-stream network, or both. Critically, even though in experiment 2.3 action-responsive ROIs are defined using the movie data, the selection criterion used - stronger responses to scenes with actions than to scenes without action - is orthogonal to the question we will test with the encoding model.

In addition to these analyses, if time allows, we will perform a whole brain analysis identifying voxels where responses are well predicted by the static stream of the model, voxels where responses are well predicted by the dynamic stream, and their overlap.

2.5 Objective 3: Multivariate Interactions between Social Perception Regions

Understanding the neural mechanism for social perception requires studying the representational content of the brain regions involved, and investigating how these regions interact as a network. Objectives 1 and

2 of this proposal focus on studying representational content. Objective 3 investigates region-to-region interactions, applying new multivariate connectivity methods [5] to the data collected in Experiments 1.2 and 2.2.

The investigation of interactions between brain regions engaged in social perception will proceed in two stages. In the first stage, we will calculate the connectivity matrix between all pairs of face-selective regions (OFA, FFA, face-pSTS) and region MT. Similarly, we will calculate a connectivity matrix between all pairs of body-selective regions (EBA, FBA, body-pSTS), and MT. In the second stage, we will test: 1) whether pSTS receives inputs only from lateral regions encoding motion information (area MT), or whether it also integrates information from ventral temporal regions (OFA); 2) whether FFA receives inputs only from ventral regions encoding motion information (OFA), or whether it also integrates information from lateral regions (area MT); and 3) whether other brain regions integrate information from the ventral (OFA and FFA) and lateral (MT and pSTS) streams. Area MT will be defined with the contrast between randomly moving point-lights and static point-lights shown in the second run of Experiments 1.1 and 2.1.

A recent technique I developed - multivariate pattern dependence (MVPD,[4, 5], see Figure 6) - models the statistical dependence between the multivariate response patterns in different brain regions. The data is divided into a training set and an independent testing set. Let $x(t)$ be the response pattern at time t in a brain region used as predictor, and $y(t)$ the response pattern in a region that is the target of prediction. The training set is used to learn a multivariate function f_{train} such that $y(t) = f_{train}(x(t)) + \epsilon(t)$ and the error $\sum_t \epsilon(t)^2$ is minimized. The predictive power of the learned function is tested using the independent testing set. Let $x_{test}(t)$ be the testing data in the predictor region and $y_{test}(t)$ the testing data in the target region. We can generate predictions $\hat{y}(t) = f_{train}(x_{test}(t))$, and calculate the proportion of the variance of y_{test} explained by $\hat{y}$ as a measure of statistical dependence. In the current implementation, f is learned using an artificial neural network with trained with stochastic gradient descent (SGD) using a mean square error loss (MSE). The code will be made available to the research community on a public repository (on Github).

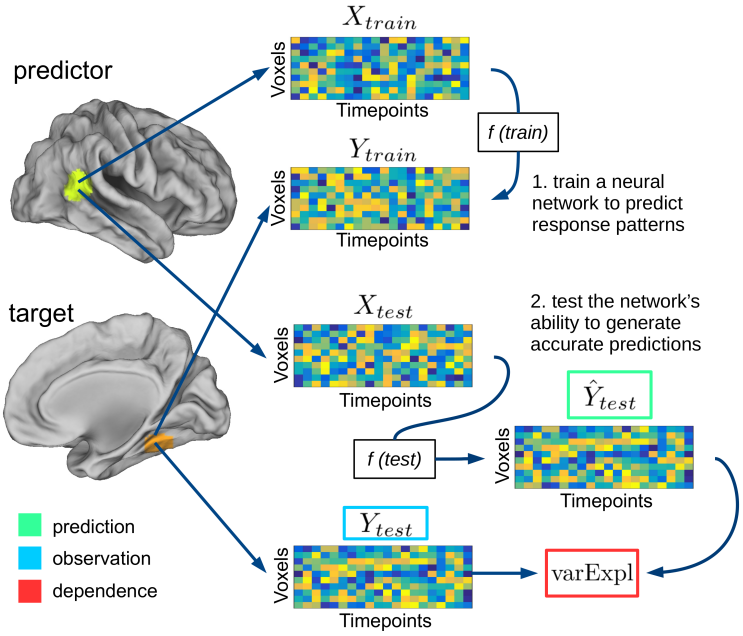

Figure 6: Multivariate pattern dependence: data analysis pipeline.

Applying MVPD to the data from Experiment 1.2, we will investigate the fusion between the lateral and ventral face-processing streams by predicting response patterns in face-selective pSTS using as inputs the response patterns in area MT, or using both MT and OFA. We will calculate whether adding OFA as a predictor increases the variance explained in the responses in face-selective pSTS. Critically, MVPD performs training and testing using independent folds of the data, ensuring that a larger number of predictors does not trivially explain more variance. Following the same logic, we will predict response patterns in face-selective FFA using as inputs response patterns in OFA, or in both OFA and MT. Finally, we will use a whole brain analysis to identify regions better predicted by response patterns in both ventral (OFA and FFA) and lateral (MT and face-selective pSTS) regions, than they are by the one of the two face-selective streams achieving highest prediction accuracy in isolation.

The same analysis procedure will be applied to the data from Experiment 2.2 to test whether responses in body-selective pSTS are better explained by response patterns in MT and EBA jointly than by MT alone, and whether responsens in FBA are better predicted by response patterns in EBA and MT jointly than by EBA alone. Finally, a whole-brain analysis will be used to identify regions better predicted by the ventral and lateral body-selective streams jointly, than by the stream achieving highest accuracy in isolation (Figure 7). Significance will be assessed with SnPM.

In the proposed project, even negative findings can provide important insights into social perception. Our Preliminary Results show that training artificial neural networks to label facial expressions from static face images leads to the spontaneous development of identity representations. If we did not find this effect in the dynamic stream of the deep network, this would suggest that dynamic and static properties of faces pose profoundly different challeges for recognition, that call for different computational solutions. Similarly, if we did not find complementarity between representations of bodies and actions, this would suggest that recognition of faces and bodies are very different problems from a computational perspective.

In the analyses of neural responses, if both the ventral and lateral streams in the human brain were predicted equally well by the static and dynamic streams in the models, these findings would challenge the view that different streams in humans are organized by whether they process static or dynamic properties - this would call for the development of new theories to account for the distinct functional roles of ventral temporal cortex and lateral temporal cortex in social perception.

In the end, in the MVPD analyses, if pSTS responses were not better explained by adding OFA responses as predictor, this would indicate that the ventral and lateral streams are more compartmentalized than we have hypothesized.

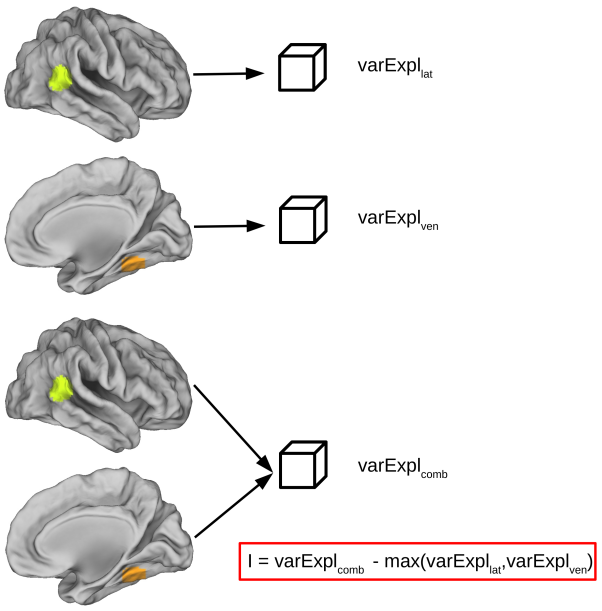

Figure 7: For the whole-brain analysis of stream fusion, we will predict each voxel's responses from multivariate response patterns in the lateral stream only (top), in the ventral stream only (middle), or in both (bottom). We will then calculate the difference between the proportion of variance explained using both streams and the proportion of variance explained using the best of the two streams taken in isolation.

2.6 Expected Outcomes and Future Plans

We expect to find that identity and expressions are represented within the same regions because of computational constraints - specifically, that recognition of expressions and identity are complementary so that learning representations for expression recognition leads to the spontaneous emergence of representations that contribute to identity recognition. Additionally, we expect that representations learned by a two-stream deep network for the recognition of dynamic faces predict neural responses in humans, and more specificlly that the static stream of the network better predicts responses in ventral temporal regions (OFA, FFA), while the dynamic stream better predict responses in lateral regions (face-selective pSTS). In the end, we expect to find that the ventral and lateral face-processing streams interact, and that face-selective pSTS integrates information from both lateral and ventral regions.

We anticipate that this architecture is not restricted to faces alone, but rather extends to social perception more generally. Thus, we expect that like representations of face identity and expressions, representations of actions and body identity are also complementary and represented within the same brain regions. We also expect that representations learned by a two-stream deep network for the recognition of actions predict neural responses in humans, so that the static stream of the network better predicts responses in ventral temporal body-selective regions (EBA, FBA), while the dynamic stream better predict responses in lateral regions (body-selective pSTS). As for face regions, we predict that body-selective pSTS integrates information from both lateral and ventral regions.

In future work beyond the scope of this proposal, we plan to test whether representations of expressions arise spontaneously when training networks to recognize face identity, and whether representations of actions arise when training networks to recognize body identity. Additionally, we plan to investigate whether networks trained with multitask training yield features that better predict neural responses. Comparing how well fMRI responses are predicted by encoding models using features from networks trained to label identity only, expression only, or both, can provide additional evidence to elucidate the role of the type of properties (static vs optic flow) vs the type of content (expression vs identity) to account for the differentiation between

the ventral and lateral pathways.

We also plan to use the two-stream neural networks to predict ECoG data and electrophysiological recordings in primates, through collaborations with other laboratories. As a first step in this direction, we are currently in the process of collaborating with another laboratory that has collected a large dataset of ECoG responses to images of facial expressions, to analyze the data using features from our single-stream model.

In the longer term, we plan to study the interconnected problems of how humans predict other people's actions, and of how perceptual representations are used as a foundation to acquire person knowledge. Understanding others' states, traits, beliefs and goals is essential for action prediction, and we hope that the challenge of action prediction can play for the study of social cognition the role that object recognition has played for the study of vision.

Other researchers will be able to use our results as a starting point to investigate how representations of faces, bodies, expression and actions are integrated with representations of objects and scenes to understand complex events. By clarifying what visual properties are represented in pSTS, our findings can also serve as a foundation to study the computational mechanisms by which visual and auditory representations are integrated in this region (see [3]), and could be used to formulate hypotheses about why audio-visual integration seems to occur in the lateral pathway rather than in the ventral pathway. Finally, our result on the representations of faces, expressions, bodies, and actions, could be used as a baseline to investigate whether there are specialized processes for the recognition of interactions between multiple agents.

2.7 Expected outcomes and future plans

3 Broader Impacts

The proposed project will have an impact on education as well as basic research, and will serve as a stepping stone for understanding impairments of social perception. With respect to education, I plan to 1) offer undergraduate and graduate students interdisciplinary training in Cognitive Neuroscience and Artificial Intelligence, so they can take advantage of the growing interactions between these two fields; 2) motivate undergraduate and graduate students to pursue careers in science; and 3) communicate the transformative potential of the interplay between Cognitive Neuroscience and AI to the general public. To achieve these goals, I will propose three initiatives: 1) enrich the research-oriented educational offerings in the curriculum through the development of new courses and workshops based on active learning, that span how Cognitive Science and Neuroscience inspired AI, the mathematical theory and the practical implementation of AI algorithms, and how AI can be used for research in Cognitive Neuroscience; 2) engage undergraduate and graduate students from diverse backgrounds with questions at the frontier of research through research projects and the interaction with leading experts in the field; and 3) organize outreach events for the community in Boston and Newton.

3.1 Enhancing the curriculum with courses at the intersection of Cognitive Neuroscience and AI

I aim to offer new courses and workshops that provide students with the knowledge foundations to understand current AI techniques, the programming skills needed to apply them to Cognitive Neuroscience research, and the critical thinking needed to interpret the implications of the results. Neuroscience and AI are usually taught in different courses offered by different departments, therefore students do not currently have the opportunity to learn about how these disciplines can inspire each other, and how to use AI for Cognitive Neuroscience research. Interdisciplinary education is critical to prepare students for scientific research, and more broadly for employment in the knowledge economy [35].

To be successful, interdisciplinary education 'must develop student capacities to integrate or synthesize disciplinary knowledge and modes of thinking' [58]. To achieve this goal, it is essential that students are motivated to acquire new knowledge and skills in a different discipline. A recent meta-analysis demonstrated the effectiveness of motivation interventions across a variety of studies taking different approaches [44] To motivate students, I will introduce them to recent examples of applications of AI to a variety of problems, with a focus on recent neuroscience research. I have taken this approach when introducing computational methods in a course on social cognition I taught in the Fall semester of 2018, and it proved effective to

spark the students' interest - they went from being skeptical of computational methods to actively asking for resources to learn more.

Proposed work: In a new course, I will teach a lecture on the potential of AI for neuroscience and more broadly for society, showing examples of applications. Next, I will teach a lecture on the mathematical foundations of deep learning, using visualizations to explain gradient descent, and ensuring that students from all background can understand the concepts. Visualizations are widely regarded as effective for education [55, 64]. Furthermore, I have used this approach to teach deep learning to two undergraduate students and a lab coordinator doing research in my group, and it has worked well in all three cases.

In my experience students acquire knowledge and skills more effectively when they participate actively to solving a problem. Recent meta-analyses indicate that active learning is particularly effective, leading to higher performance in examinations and lower odds of failing [22]. For this reason, after the first two lectures, I will ask students to choose a project, among three options selected to be both challenging, feasible, and relevant to Cognitive Neuroscience. Letting students choose the project will help to give them a sense of ownership of the problem they need to solve. Students will be divided into small groups.

Previous research indicates that work in small groups can result in higher achievement [36]. However, several factors are needed to obtain this outcome. First, group members must perceive that they cannot succeed unless all group members succeed [36]. To this end, students will be told that each group member is responsible for the overall results. Another key factor for the success of work in groups is 'promotive interaction' [36], in which group members facilitate each other's efforts for the project. To the extent possible given the class composition, I will organize groups so that each group includes students from different backgrounds, who can learn from each other, encouraging promotive interaction. Finally, each group will have to decide what strategies to use in their project, choosing among many possible options. This will lead group members to practice communication, conflict resolution, and leadership skills. The course will require as personnel the PI (me), and will require budget for the purchase of 2 GPU workstations to be used for education purposes.

In addition to the course, I will organize a graduate workshop on AI and neuroscience, in which graduate students and advanced undergraduates working in the lab will present ongoing neuroscience research that uses AI techniques. Before presenting their research project, students will explain a key technique used in the project in a way that is accessible to a broad audience of Cognitive Neuroscience researchers. The workshop will be open to other laboratories in the department, and other students will be encouraged to attend, to foster the development of collaborations and introduce the graduate student community in the department to the use of AI techniques for neuroscience research. The workshop will require the organization and supervision of the PI (me), and the participation of the members of my laboratory. Together, the proposed course and workshop will help **train a new generation of Cognitive Neuroscientists with advanced Artificial Intelligence skills**.

Evaluation: In the course there will be two evaluation sessions, one in the middle of the semester, and one at the end. At each evaluation session, I will ask each group to present their work to the class, assigning a part of the presentation to each student. Additionally, each group will write a report. Each group member will be assigned a part of the report. The group's performance will be evaluated as a whole. Thanks to the evaluation session in the middle of the semester, students in each group will have the opportunity to learn from the other groups ('peer learning', [9]), and will have time to incorporate the things they learned into their projects in the remaining part of the course. The workshop will be assessed using 1) a survey given to the audience asking whether the presentations were clear and useful, and 2) an evaluation after one year of how many new projects in the department are using AI to investigate neuroscience questions.

3.2 Engaging students from diverse backgrounds with questions at the frontier of research

Boston College has 27% minority enrollment and 55% women, providing an ideal context to engage students from diverse backgrounds in research and to encourage them to pursue careers in science. In addition to the standard research opportunities available in the lab, I propose two new activities to involve students in cutting edge interdisciplinary research, together with a strategy to engage a diverse population of students.

Proposed work: Summer is an ideal time of the year for undergraduates to gain research experience free of the demands of coursework. During my time as a postdoc and as a faculty, I have mentored several

undergraduate students for summer research projects. In particular, as a postdoc at MIT I had the opportunity to mentor an african-american student for the MSRP (MIT Summer Research Program) - this was a transformative experience, and he went on to pursue an MD at Yale. Motivated by this experience, I propose to establish a summer research opportunity in my laboratory for undergraduate students from underrepresented groups. Students will receive a summer salary to support their permanence on campus. For recruitment, we will seek the help of the Thea Bowman AHANA (African, Hispanic, Asian and Native American) and Intercultural Center and of student associations such as the Undergraduate Government of Boston College (UGBC)'s Diversity and Inclusion division.

Inspiration is critical to encourage students to pursue careers in science. Often, inspiration comes from interacting with people who can push the boundaries of what we think is possible. I propose to offer students the opportunity to interact directly with leading researchers working in fields at the intersection of Cognitive Neuroscience and AI, organizing a bi-annual conference at Boston College. The conference will consists of a series of talks as well as poster sessions in which students will have the opportunity to discuss their projects with experts in the field (developed in the course described in Aim 1 or in laboratory research). In addition, we will organize a faculty-student lunch in which students will have the opportunity to ask researchers questions about their work and about their experiences in academic careers. Research suggests that students tend to choose career role models with the same race [38] - in our conference we will strive to invite a diverse group of speakers. Each of the two planned conferences will require a budget of $8000 to cover travel costs for the speakers, rental fees for a conference room, and refreshments.

The summer program and conferences will **inspire a diverse population of students and expose them to the world of research**.

Evaluation: At the end of the summer program, students will deliver a presentation to the rest of the laboratory, and prepare a written report. These activities will serve as learning experiences, and will be used to assess the student's work during the summer. In addition, we will ask students to report what program or job they enroll in after they complete their undergraduate degree, to evaluate the impact of the program on career choices. We will make clear that reporting this information is on a voluntary basis. Evaluation of the workshop will be twofold: we will ask students to rate their experience with a survey, and we will ask the invited speakers to evaluate the workshop and to offer suggestion for the following editions.

3.3 Organizing outreach events for the community in Boston and Newton

Outreach to the general public is critical for the future of science. Communicating the importance and implications of research is necessary to motivate support for science. Furthermore, new generations reflect on career opportunities in the context of their families [19], and outreach events for the general public offer an ideal opportunity to stimulate conversation involving children as well as their parents. In the April 2019, two of my students presented at the talk series 'Artificial intelligence meets neuroscience' organized as part of the Cambridge Science festival for the general public. Several families attended the event and asked questions about careers in science. Boston College, with its location west of the city, proximity to the I-90 highway, and easy access to parking, can be an ideal meeting place for broader swaths of the MA population, including those who live in Central and Western Massachusetts areas (i.e. Worcester, Springfield) and who have relatively limited access to such outreach events.

Proposed work: I plan to organize a bi-annual event at Boston College to introduce research at the intersection of Cognitive Neuroscience and AI to a general public. The event organized for the Cambridge Science festival included only talks, so it was engaging for adults but not as much for children. To improve on this aspect, I will organize an event that includes not only presentations, but also activities for a younger public, including demonstrations on the functional role of different brain regions and on the functioning of fMRI. I have organized similar activities as a postdoc for a 'Science on Saturday' outreach event at MIT, that were well received by the children participating. In addition, to reach a broader audience, I will organize a 3 hour IAmA event on reddit. The public event and the IAmA will **raise awareness about Cognitive Neuroscience research and its potential to contribute society**.

Evaluation: The success of the outreach event will be measured using a survey (including both ratings and open-answer questions) as well as interviews to a sample of participant. The evaluations obtained after the first event will be used as formative evaluations and will be used to improve the following event.

The workshop will require the organization and supervision of the PI (me), and the participation of the members of my laboratory. Members of other laboratories at Boston College who have since started adopting artificial intelligence methods will be encouraged to participate.

3.4 Integration of Research and Education

The proposed education plan is tightly coupled with the research program we aim to pursue. Many facets of the education plan rely on exposing students directly to the process of discovery, whether at the undergraduate or graduate level. The proposed courses are largely based on practical, hands-on activities using the very same kinds of methods, software and hardware used for research, and ongoing research efforts will be used as examples of what can be done with these methods to inspire the students. Projects tackled by students in class might draw from components of our research plan, so that students might be invited to reproduce our research findings. The workshop I plan to organize is perhaps the culmination of this integration between research and education, where students can present their work to experts in the field, can learn about state of the art research directions, and can be inspired by ongoing projects and inspire them in turn with their questions.

3.5 Impact of the education plan in a local and national context

At Boston College, **the proposed education opportunities address key student demands**. Boston College's department of Computer Science does not have a Ph.D program and has only 8 faculty members, therefore it has difficulty offering enough courses to satisfy the students' demand. Access to several Computer Science courses is restricted to students majoring in Computer Science. In this context, the course I propose to teach would fill a critical need, not only for majors in Psychology and Neuroscience, but also for other student more broadly interested in Computer Science. Additionally, it would provide support and intersect with a major expansion hiring initiative in Computer Science currently ongoing at Boston College. Importantly, the course would also be the first offering in computational Neuroscience in the department, playing a significant role in shaping the new Neuroscience major that will be launched in academic year 2019-2020.

More broadly, **education opportunities at the intersection of Neuroscience and Artificial Intelligence are growing necessity in the country**. The impressive expansion of the technology industry has led to a drain of talents from the academia [42] - for example, in 2015, Uber hired 40 AI researchers from Carnegie Mellon University. Technology companies are playing a key role to drive the economy of the United States, but the drain of talents from the academia raises the question of who will train the next generations of researchers and scientists. This problem hits the fields of Neuroscience and Cognitive Science particularly hard, at a time where technological innovation in artificial intelligence is proving increasingly vital to drive progress on research questions [41].

In parallel, many **key advances in AI have been the result of inspiration from the brain** - from pooling operations inspired to Hubel and Wiesel's pioneering work on complex cells, to the more recent cases of DenseNets [34], inspired to the dense connectivity of primate visual cortex, and of deep networks with attention [70].

The proposed project will contribute to our nation's current and future needs to train a new generation of neuroscientists who can use state of the art AI techniques, not only for pure research, but also for the diagnosis of brain cancer and psychiatric disorders. Furthermore, it will contribute to the training of future researchers that can leverage a deeper understanding of the brain to inspire progress in AI.

3.6 Impact on basic research

In addition to this project's contribution to education, the discoveries resulting from the proposed studies will be an important step forward towards understanding not only social perception, but social cognition more broadly. Observing others' actions and expressions in context is a key source of what we know about others [7]. In turn, person knowledge is necessary for our ability to interact in a social context. Thus, investigating its neural and computational mechanisms (as proposed in this project) is a **fundamental stepping stone towards understanding social cognition in general**. Understanding social cognition is central to tackle the challenges we face in the XXI century. Problems ranging from discrimination to global warming are the result of collective behavior, and call for solutions that can take advantage of a scientific understanding of how people act and interact with each other.

The proposed project is unique in that it takes as a starting point a novel finding in neuroscience - the decoding of both expressions and identity within the same brain regions [61, 3] - to generate new insights about the properties of deep networks, and in turn, it uses the new insights about deep networks to offer a computational theory of social perception that can be tested with neural measures. As such, the project is an example of how the interplay between neuroscience and artificial intelligence can be bidirectional, with each discipline offering inspiration for the other in a symbiotic interaction.

Additionally, two **techniques adopted in this project are innovative and have the potential to be applied to a variety of problems in other domains**. The first is the suggested strategy to use deep analysis-by-synthesis to generate realistic dynamic stimuli and produce a large training dataset for deep neural networks. This type of strategy can be applied to study a variety of phenomena with complex dynamics, where a generative model is available to reproduce observations at individual timepoints, but the naturalistic temporal dynamics of the observations are uncharted. The second is Multivariate Pattern Dependence (MVPD): its application in this project can serve as an example of the potential and flexibility of this approach, which can be adopted to study multivariate interactions between brain regions across different domains of cognition. The formal structure of MVPD is very general, therefore its application is not restricted to fMRI data, but can also be extended to calcium imaging and electrophysiology.

In the end, the novel findings on the mechanisms for the recognition of face identity and facial expressions stemming from the proposed studies will help to understand the impairments occurring in prosopagnosia, a disorder of face recognition which affects millions of people in the United States [62]; our findings on the computational and neural bases of the recognition of bodies and biological motion will help to understand atypical development of biological motion processing in autism spectrum disorders [1], which affects millions in the USA and has an estimated annual economic cost between $11 and $60 billions [43, 11].

Overall, the proposed project will have an important impact on education, where it will address a timely and critical need for training scientists with proficiency in artificial intelligence, on basic research, and on our nation's preparedness to tackle key challenges of the XXI century through transformative, interdisciplinary research.

4 Timeline

The proposed project will be completed according to the following timeline:

	Year 1			Year 2			Year 3			Year 4			Year 5		
	Fall	Spring	Summ.	Fall	Spring	Summ.	Fall	Spring	Summ.	Fall	Spring	Summ.	Fall	Spring	Summ.
Research Aim 1															
Generate videos: deep analysis by synthesis	■														
Train two-stream network for expressions		■	■												
Test two-stream network with face identity			■												
Implement fMRI experiments 1.1-1.2			■	■											
Collect fMRI data				■	■										
Model fMRI responses from network features						■	■								
Research Aim 2															
Train two-stream network for actions								■	■						
Test two-stream network with body identity									■						
Select stimuli for Experiment 2.2						■	■	■							
Implement fMRI Experiment 2.2									■	■					
Collect fMRI data											■	■			
Model fMRI responses from network features													■		
Research Aim 3															
Test MVPD for face regions											■	■			
Test MVPD for body regions														■	■
Education Aim 1															
Prepare course						■	■								
Teach course								■						■	
Revise course												■	■		
Neuroscience and AI graduate workshops	■			■			■			■			■		
Education Aim 2															
Undergraduate summer program			■			■			■			■			
Bi-annual conference organization					■						■				
Bi-annual conference								■						■	
Education Aim 3															
Preparation outreach activity				■						■					
Outreach: neuroscience and AI					■						■				
IAmA							■				■				

5 Results from prior NSF support

Not applicable.

References

[1] Dagmara Annaz, Anna Remington, Elizabeth Milne, Mike Coleman, Ruth Campbell, Michael SC Thomas, and John Swettenham. Development of motion processing in children with autism. *Developmental science*, 13(6):826–838, 2010.

[2] Stefano Anzellotti and Alfonso Caramazza. From parts to identity: invariance and sensitivity of face representations to different face halves. *Cerebral Cortex*, 26(5):1900–1909, 2015.

[3] Stefano Anzellotti and Alfonso Caramazza. Multimodal representations of person identity individuated with fmri. *Cortex*, 89:85–97, 2017.

[4] Stefano Anzellotti, Alfonso Caramazza, and Rebecca Saxe. Multivariate pattern dependence. *PLoS computational biology*, 13(11):e1005799, 2017.

[5] Stefano Anzellotti and Marc N Coutanche. Beyond functional connectivity: investigating networks of multivariate representations. *Trends in cognitive sciences*, 22(3):258–269, 2018.

[6] Stefano Anzellotti, Scott L Fairhall, and Alfonso Caramazza. Decoding representations of face identity that are tolerant to rotation. *Cerebral Cortex*, 24(8):1988–1995, 2013.

[7] Stefano Anzellotti and Liane L Young. The acquisition of person knowledge. 2019.

[8] Michael S Beauchamp, Kathryn E Lee, James V Haxby, and Alex Martin. Fmri responses to video and point-light displays of moving humans and manipulable objects. *Journal of cognitive neuroscience*, 15(7):991–1001, 2003.

[9] David Boud, Ruth Cohen, and Jane Sampson. *Peer learning in higher education: Learning from and with each other*. Routledge, 2014.

[10] Vicki Bruce and Andy Young. Understanding face recognition. *British journal of psychology*, 77(3):305–327, 1986.

[11] Ariane VS Buescher, Zuleyha Cidav, Martin Knapp, and David S Mandell. Costs of autism spectrum disorders in the united kingdom and the united states. *JAMA pediatrics*, 168(8):721–728, 2014.

[12] Beatriz Calvo-Merino, Daniel E Glaser, Julie Grèzes, Richard E Passingham, and Patrick Haggard. Action observation and acquired motor skills: an fmri study with expert dancers. *Cerebral cortex*, 15(8):1243–1249, 2004.

[13] Joao Carreira and Andrew Zisserman. Quo vadis, action recognition? a new model and the kinetics dataset. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6299–6308, 2017.

[14] Cristiana Cavina-Pratesi, Jason D Connolly, Simona Monaco, Teresa D Figley, A David Milner, Thomas Schenk, and Jody C Culham. Human neuroimaging reveals the subcomponents of grasping, reaching and pointing actions. *Cortex*, 98:128–148, 2018.

[15] Katharina Dobs, Wei Ji Ma, and Leila Reddy. Near-optimal integration of facial form and motion. *Scientific reports*, 7(1):11002, 2017.

[16] Katharina Dobs, Johannes Schultz, Isabelle Bülthoff, and Justin L Gardner. Task-dependent enhancement of facial expression and identity representations in human cortex. *Neuroimage*, 172:689–702, 2018.

[17] Paul E Downing, Yuhong Jiang, Miles Shuman, and Nancy Kanwisher. A cortical area selective for visual processing of the human body. *Science*, 293(5539):2470–2473, 2001.

[18] Oscar Esteban, Christopher J Markiewicz, Ross W Blair, Craig A Moodie, A Ilkay Isik, Asier Erramuzpe, James D Kent, Mathias Goncalves, Elizabeth DuPre, Madeleine Snyder, Hiroyuki Oya, Satrajit S. Ghosh, Jessey Wright, Joke Durnez, Russell A. Poldrack, and Krzysztof J. Gorgolewsky. Fmriprep: a robust preprocessing pipeline for functional mri. *Nature methods*, 16(1):111, 2019.

[19] Tamara R Ferry, Nadya A Fouad, and Philip L Smith. The role of family context in a social cognitive model for career-related choice behavior: A math and science perspective. *Journal of Vocational Behavior*, 57(3):348–364, 2000.

[20] Christopher J Fox, Hashim M Hanif, Giuseppe Iaria, Bradley C Duchaine, and Jason JS Barton. Perceptual and anatomic patterns of selective deficits in facial identity and expression processing. *Neuropsychologia*, 49(12):3188–3200, 2011.

[21] Christopher J Fox, Giuseppe Iaria, and Jason JS Barton. Defining the face processing network: optimization of the functional localizer in fmri. *Human brain mapping*, 30(5):1637–1651, 2009.

[22] Scott Freeman, Sarah L Eddy, Miles McDonough, Michelle K Smith, Nnadozie Okoroafor, Hannah Jordt, and Mary Pat Wenderoth. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23):8410–8415, 2014.

[23] Winrich Freiwald, Bradley Duchaine, and Galit Yovel. Face processing systems: from neurons to real-world social perception. *Annual Review of Neuroscience*, 39:325–346, 2016.

[24] Ricardo Gattass and Charles G Gross. Visual topography of striate projection zone (mt) in posterior superior temporal sulcus of the macaque. *Journal of Neurophysiology*, 46(3):621–638, 1981.

[25] Michaël Gilbert, Samuel Demarchi, and Isabel Urdapilleta. Facshuman a software to create experimental material by modeling 3d facial expression. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, pages 333–334. ACM, 2018.

[26] Nikolaos Gkalelis, Hansung Kim, Adrian Hilton, Nikos Nikolaidis, and Ioannis Pitas. The i3dpost multi-view and 3d human action/interaction database. In *2009 Conference for Visual Media Production*, pages 159–168. IEEE, 2009.

[27] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[28] Ian J Goodfellow, Dumitru Erhan, Pierre Luc Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski, Yichuan Tang, David Thaler, Dong-Hyun Lee, Yingbo Zhou, Chetan Ramaiah, Fangxiang Feng, Ruifan Li, Xiaojie Wang, Dimitris Athanasakis, John Shawe-Taylor, Maxim Milakov, John Park, Radu Ionescu, Marius Popescu, Cristian Grozea Grozea, James Bergstra, Jingjing Xie, Lukasz Romaszko, Bing Xu, Zhang Chuang, and Yoshua Bengio. Challenges in representation learning: A report on three machine learning contests. In *International Conference on Neural Information Processing*, pages 117–124. Springer, 2013.

[29] Kalanit Grill-Spector, Nicholas Knouf, and Nancy Kanwisher. The fusiform face area subserves face perception, not generic within-category identification. *Nature neuroscience*, 7(5):555, 2004.

[30] Alon Hafri, John C Trueswell, and Russell A Epstein. Neural representations of observed actions generalize across static and dynamic visual input. *Journal of Neuroscience*, 37(11):3056–3071, 2017.

[31] Michael Hanke, Nico Adelhöfer, Daniel Kottke, Vittorio Iacovella, Ayan Sengupta, Falko R Kaule, Roland Nigbur, Alexander Q Waite, Florian Baumgartner, and Jörg Stadler. A studyforrest extension, simultaneous fmri and eye gaze recordings during prolonged natural stimulation. *Scientific data*, 3:160092, 2016.

[32] Bashar Awwad Shiekh Hasan, Mitchell Valdes-Sosa, Joachim Gross, and Pascal Belin. "hearing faces and seeing voices": Amodal coding of person identity in the human brain. *Scientific reports*, 6:37494, 2016.

[33] James V Haxby, Elizabeth A Hoffman, and M Ida Gobbini. The distributed human neural system for face perception. *Trends in cognitive sciences*, 4(6):223–233, 2000.

[34] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[35] W Jacob. Interdisciplinary trends in higher education. 2015.

[36] Roger T Johnson and David W Johnson. Active learning: Cooperation in the classroom. *The annual report of educational psychology in Japan*, 47:29–30, 2008.

[37] Nancy Kanwisher, Josh McDermott, and Marvin M Chun. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience*, 17(11):4302–4311, 1997.

[38] Danesh Karunanayake and Margaret M Nauta. The relationship between race and students' identified career role models and perceived role model influence. *The Career Development Quarterly*, 52(3):225–234, 2004.

[39] Will Kay, Joao Carreira, Karen Simonyan, Brian Zhang, Chloe Hillier, Sudheendra Vijayanarasimhan, Fabio Viola, Tim Green, Trevor Back, Paul Natsev, Mustafa Suleyman Suleyman, and Andrew Zisserman. The kinetics human action video dataset. *arXiv preprint arXiv:1705.06950*, 2017.

[40] Dorit Kliemann, Hilary Richardson, Stefano Anzellotti, Dima Ayyash, Amanda J Haskins, John DE Gabrieli, and Rebecca R Saxe. Cortical responses to dynamic emotional facial expressions generalize across stimuli, and are sensitive to task-relevance, in adults with and without autism. *Cortex*, 103:24–43, 2018.

[41] Nikolaus Kriegeskorte. Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annual review of vision science*, 1:417–446, 2015.

[42] Lars Kunze. Can we stop the academic ai brain drain?, 2019.

[43] Tara A Lavelle, Milton C Weinstein, Joseph P Newhouse, Kerim Munir, Karen A Kuhlthau, and Lisa A Prosser. Economic burden of childhood autism spectrum disorders. *Pediatrics*, 133(3):e520–e529, 2014.

[44] Rory A Lazowski and Chris S Hulleman. Motivation interventions in education: A meta-analytic review. *Review of Educational research*, 86(2):602–640, 2016.

[45] Yichen Li, Rebecca Saxe, and Stefano Anzellotti. Intersubject mvpd: Empirical comparison of fmri denoising methods for connectivity analysis. *BioRxiv*, page 456970, 2018.

[46] Angelika Lingnau and Paul E Downing. The lateral occipitotemporal cortex in action. *Trends in cognitive sciences*, 19(5):268–277, 2015.

[47] Patrick Lucey, Jeffrey F Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pages 94–101. IEEE, 2010.

[48] Jin Narumoto, Tomohisa Okada, Norihiro Sadato, Kenji Fukui, and Yoshiharu Yonekura. Attention to emotion modulates fmri activity in human right superior temporal sulcus. *Cognitive Brain Research*, 12(2):225–231, 2001.

[49] Thomas Naselaris and Kendrick N Kay. Resolving ambiguities of mvpa using explicit models of representation. *Trends in cognitive sciences*, 19(10):551–554, 2015.

[50] Thomas Naselaris, Kendrick N Kay, Shinji Nishimoto, and Jack L Gallant. Encoding and decoding in fmri. *Neuroimage*, 56(2):400–410, 2011.

[51] Kathryn C O'Nell, Rebecca Saxe, and Stefano Anzellotti. Recognition of identity and expressions as integrated processes. 2019.

[52] Alice J O'Toole, P Jonathon Phillips, Samuel Weimer, Dana A Roark, Julianne Ayyad, Robert Barwick, and Joseph Dunlop. Recognizing people from dynamic and static faces and bodies: Dissecting identity with a fusion approach. *Vision research*, 51(1):74–83, 2011.

[53] Maja Pantic, Michel Valstar, Ron Rademaker, and Ludo Maat. Web-based database for facial expression analysis. In *2005 IEEE international conference on multimedia and Expo*, pages 5–pp. IEEE, 2005.

[54] Marius V Peelen, Anthony P Atkinson, and Patrik Vuilleumier. Supramodal representations of perceived emotions in the human brain. *Journal of Neuroscience*, 30(30):10127–10134, 2010.

[55] Linda M Phillips, Stephen P Norris, and John S Macnab. *Visualization in mathematics, reading and science education*, volume 5. Springer Science & Business Media, 2010.

[56] David Pitcher, Daniel D Dilks, Rebecca R Saxe, Christina Triantafyllou, and Nancy Kanwisher. Differential selectivity for dynamic versus static information in face-selective cortical regions. *Neuroimage*, 56(4):2356–2363, 2011.

[57] David Pitcher, Bradley Duchaine, and Vincent Walsh. Combined tms and fmri reveal dissociable cortical pathways for dynamic and static face perception. *Current Biology*, 24(17):2066–2070, 2014.

[58] Diana Rhoten, V Boix Mansilla, Marc Chun, and Julie Thompson Klein. Interdisciplinary education at liberal arts institutions. *Teagle Foundation White Paper. Retrieved June*, 13:2007, 2006.

[59] Rebecca F Schwarzlose, Chris I Baker, and Nancy Kanwisher. Separate face and body selectivity on the fusiform gyrus. *Journal of Neuroscience*, 25(47):11055–11059, 2005.

[60] Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. In *Advances in neural information processing systems*, pages 568–576, 2014.

[61] Amy E Skerry and Rebecca Saxe. A common neural code for perceived and inferred emotion. *Journal of Neuroscience*, 34(48):15997–16008, 2014.

[62] Tirta Susilo and Bradley Duchaine. Advances in developmental prosopagnosia research. *Current opinion in neurobiology*, 23(3):423–429, 2013.

[63] Lucia M Vaina, Jeffrey Solomon, Sanjida Chowdhury, Pawan Sinha, and John W Belliveau. Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Sciences*, 98(20):11656–11661, 2001.

[64] Karen L Vavra, Vera Janjic-Watrich, Karen Loerke, Linda M Phillips, Stephen P Norris, and John Macnab. Visualization in science education. *Alberta Science Education Journal*, 41(1):22–30, 2011.

[65] Mark W Woolrich, Brian D Ripley, Michael Brady, and Stephen M Smith. Temporal autocorrelation in univariate linear modeling of fmri data. *Neuroimage*, 14(6):1370–1386, 2001.

[66] Moritz F Wurm and Angelika Lingnau. Decoding actions at different levels of abstraction. *Journal of Neuroscience*, 35(20):7727–7735, 2015.

[67] Daniel LK Yamins and James J DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, 19(3):356, 2016.

[68] Ilker Yildirim, Mario Belledonne, Winrich Freiwald, and Joshua Tenenbaum. Efficient inverse graphics in biological face processing. *bioRxiv*, page 282798, 2019.

[69] Galit Yovel and Alice J O'Toole. Recognizing people in motion. *Trends in Cognitive Sciences*, 20(5):383–395, 2016.

[70] Dongbin Zhao, Yaran Chen, and Le Lv. Deep reinforcement learning with visual attention for vehicle classification. *IEEE Transactions on Cognitive and Developmental Systems*, 9(4):356–367, 2016.

[71] Guoying Zhao, Xiaohua Huang, Matti Taini, Stan Z Li, and Matti PietikäInen. Facial expression recognition from near-infrared videos. *Image and Vision Computing*, 29(9):607–619, 2011.

[72] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320, 2005.