Multivariate Neuroimaging in Social and Personality Psychology

Robert S. Chavez[1], **William A. Cunningham**[2], **and Elliot T. Berkman**[1]

[1]University of Oregon
[2]University of Toronto

Methodological approaches in Social Neuroscience have been rapidly evolving in recent years. Fueling these changes is the adoption of a variety of multivariate approaches that allow researchers to ask a wider and richer set of questions than previously possible with standard univariate methods. In this chapter, we introduce several of the most popular multivariate methods and discuss how they can be used to advance our understanding of how social cognition and personality processes are represented in the brain. These methods have the potential to allow neuroscience measures to inform and advance theories in Social and Personality Psychology more directly and are likely to become the dominant approaches in Social Neuroscience in the near future.

Note: This is a forthcoming chapter for the *Handbook of Research Methods in Social and Personality Psychology*. The final version may differ from the version here.

Neuroscience methods have been a part of social and personality psychology for several decades (Cacioppo & Berntson, 1992). In this time, the once nascent field of social neuroscience has grown from niche interest to a rapidly maturing research area, spawning its own theoretical contributions and debates. By now, many social and personality psychologists have a basic familiarity with how neuroimaging methods and approaches are used in social neuroscience. However, many of these researchers may also be less aware of some of the major shifts and advancements in multivariate approaches—pioneered for functional magnetic resonance imaging (fMRI)—in social neuroscience that allow researchers to ask a much broader and richer set of questions than previously possible.

Early fMRI work in social neuroscience borrowed from standard methodological approaches in cognitive neuroscience. For example, to determine whether there are face specific regions of the brain, researchers might show pictures of faces and other stimuli to participants during fMRI scanning. In each part of the brain, researchers would fit a model to estimate a response to each type of stimulus (e.g., a face or a house). This would be done across every voxel—a volumetric pixel at one place in the brain that used as the unit of measurement in neuroimaging—one at a time. These responses would then be contrasted against the other stimuli at each voxel to test where there was a greater response to one condition versus another. This process would be repeated, one-by-one, for every voxel throughout the brain. In the example here, researchers have consistently found that one area, the Fusiform Face Area (FFA) tends to show the most activation to faces, whereas another, the Parihippocampal Place Area (PPA) tends to show the most activation to houses/places. To make this discovery each and every voxel in the brain needed to be compared (upwards of tens of thousands of brain voxels) to determine which were more face specific, which were more place specific, and which did not seem to show a preference for one over the other. Given the number of comparisons, researchers would often use sophisticated clustering methods (a voxel would be deemed significant only if it reached a conservative significant threshold and was surrounded by other voxels that also were significant), which would then be applied to ensure statistical robustness across these multiple comparisons. A description of these methods accessible to social and personality psychologists and a summary of their primary strengths and limitations can be found in the previous edition of his Handbook (Berkman et al., 2014). Many of these methods are critical and remain the benchmark of how we estimate brain responses at each voxel of the brain.

However, we also know that parts of the brain within these voxels do not work in isolation. The brain is a massively intricate parallel processing system, and it is a truism that complex cognitive and behavioral phenomena almost never have a simple one-to-one mapping with individual regions of the brain (Cacioppo et al., 2003). For example, it could be that the same brain region as a whole is active for two tasks, but it might be doing two very different things. Analytic tools need to incorporate brain responses from different areas simultaneously to more realistically model how information is represented across systems of the brain. To meet the challenge of understanding how psychological processes are represented in the brain, researchers have developed a suite of quantitative approaches that leverage the multivariate nature of brain imaging data. These methods do not consider each voxel in isolation from one another, but rather assume that brain function can be identified by looking for patterns of activation – activation in voxel 1 may mean something quite different when paired with activation in voxel 2 versus voxel 3. Multivariate approaches have become central to contemporary social neuroscience and have shaped the kinds of questions that can be asked and inferences that are afforded. As such, the focus of this chapter is on recent advancements in multivariate neuroimaging methods and how they have been utilized to inform social and personality psychology. The bulk of these advancements augment traditional approaches to neuroimaging rather than replace them. Contemporary research deploys a range of methods, including a mixture of univariate and multivariate tools, depending on their ability to answer the research question at hand. Furthermore, although this chapter will mostly focus on fMRI methods, these multivariate approaches can, in principle, be applied to any neuroimaging or psychophysiological modality where simultaneous recordings are captured from multiple sites. (The use of multivariate approaches in non-MRI research will be highlighted briefly later.) The goal of this chapter is to introduce these contemporary methodological approaches in social neuroscience and to refashion readers' views of the role of neuroimaging in social and personality psychology.

History & Utility of Multivariate Neuroimaging

Faces of Change.

Social neuroscience employs several different techniques to characterize how psychological phenomena are reflected in the brain. For years, techniques such as electroencephalography (measuring scalp electrical current as an index of neural activity) and positron emission tomography (measuring the slow decay of radioactive tracers injected into a participant) were the most commonly used modalities in human neuroscience. By the early 2000s, however, fMRI had come to dominate the landscape of high-profile studies of the brain basis of social cognition. This method had better spatial resolution than electroencephalography and better temporal resolution than positron emission tomography and sat as a compromise between the two. Traditional approaches to fMRI analysis—referred to as univariate approaches—considered each part of the brain individually. As the face example above illustrated, researchers would measure brain responses to a particular stimulus category within individual voxels and contrast those with responses to a different category of stimuli (faces vs houses; White-American vs Black American faces; stimuli associated with monetary rewards or foot shocks). If the difference in responses is large enough in a voxel, one can infer that there was greater activation in a region for one condition than another. When done with well-controlled conditions, this approach is a useful technique for establishing the location of what parts of the brain are differentially responsive when people engage in social cognition or processing socially relevant information.

One of the most immediate and important sources of social information is the face. In the early years of fMRI, there was much interest in the neural basis of facial processing. In a landmark study, Kanwisher et al., (1997) demonstrated that a portion of the right ventral temporal lobe called the fusiform gyrus was reliably and robustly active to faces more than any other visual object category using standard univariate methods. This region was dubbed the fusiform face area (FFA) and early studies claimed that this was a category-specific brain module for facial processing (Kanwisher et al., 1997). However, the initial interpretations of these findings were challenged a few years later. In a seminal paper, Haxby et al., (2001) used a multivariate approach and found that, although the FFA was more responsive to faces than other regions were, distributed patterns of activity within the FFA could accurately dissociate multiple object categories, meaning it can distinguish categories from each other on the basis of their activity pattern. In other words, there was more activation for faces compared to other objects when blurring across all voxels within these regions - a quantitative difference. But, these regions showed unique patterns when examining all the individual voxels – a qualitative distinction. This result suggests that although the FFA shows the strongest activity to faces per the univariate analysis, general object categorization is also represented in patterns of activity that are distributed across voxels in FFA and can be captured by these multivariate methods. Despite having seemingly contradictory interpretations of the FFA data, both findings have been replicated dozens of times and are well established effects (Duchaine & Yovel, 2015).

The contrast between the results from univariate and multivariate approaches simply indicates that the classes of methods address different questions (Jimura & Poldrack, 2012). Classic univariate methods, such as the ones deployed by Kanwisher and colleagues, are based on identifying regions where there are differences in the magnitude of the response between stimulus classes. Thus, if the research question is "what part of the brain responds more to faces than other categories?" univariate methods are powerful and appropriate. Multivariate methods—referred to as multivariate pattern analysis (MVPA)—answer different questions. Rather than asking where there is a greater response to a stimulus, MVPA asks whether information related to a psychological process can be statistically decoded, or represented by a weighted combination of voxel activity, within distributed patterns of activity in multiple voxels of a given region irrespective of the magnitude of its average response. For example, a portion of the dorsal anterior cingulate cortex is implicated in processing both physical pain and social pain (Eisenberger, Lieberman, & Williams, 2003). However, Woo et al. (2014) used MVPA to show that these two conditions are reliably distinguished based on the patterns of voxel activation, even when they have similar activation across all the voxels on average. Similarly, returning to the domain of face processing, although FFA has a greater average response to faces than other categories, MVPA methods can accurately detect the percept of other object categories within the same region due to their distributed patterns (see: Figure 1). Haxby et al. (2001) was not the first paper to deploy multivariate methods in fMRI (see: McIntosh et al., 1996) but that paper is often cited as the one that pioneered MVPA and was the basis for many of the advancements that followed. Since this early work, a variety of MVPA methods have been developed to answer different kinds of questions.

Rethinking Neuroimaging for Social & Personality Psychology.

Though the focus away from magnitude differences to distributed activity patterns may seem like a subtle shift, it ac-

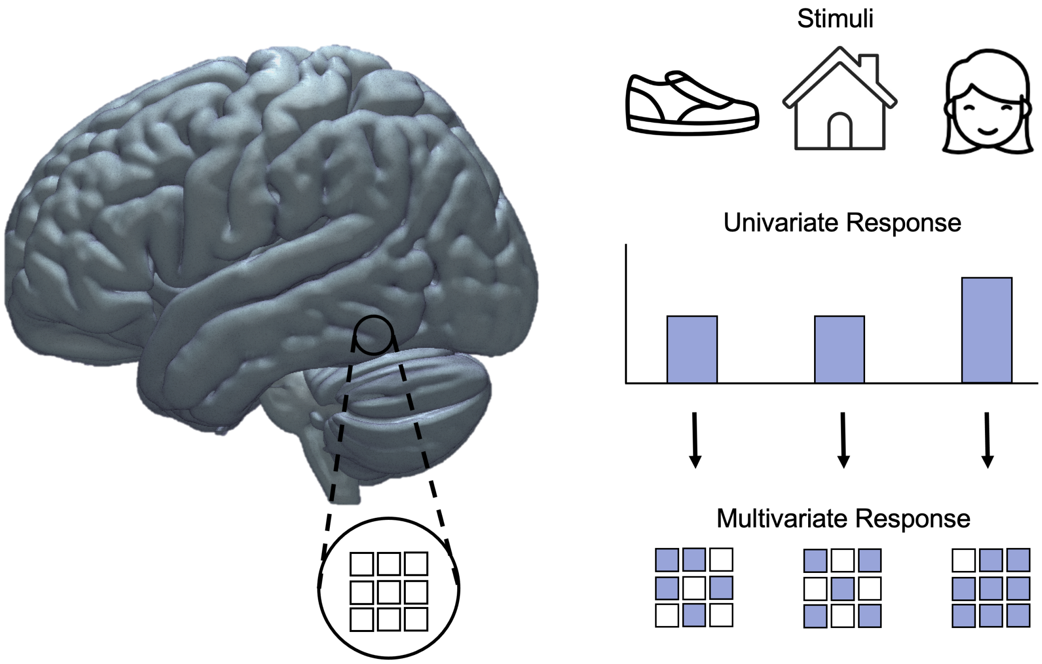

Fig. 1. *Schematic of Univariate and Multivariate Brain Responses.*The basic unit of measurement of the brain using neuroimaging are volumetric pixels called voxels. When measuring their response to different stimulus categories (e.g., shoes, houses, or faces), response activity is measured in each voxel of a given region. In univariate analyses, activity is averaged across all the voxels within the region to test whether there is a greater response for one category versus another in that region. In multivariate analyses, voxels are not averaged or aggregated but instead used in a multi-variable framework. This schematic shows an example of a case in which there is no average univariate difference between shoe and house categories but markedly different multivariate patterns of response while maintaining both multivariate and univariate differences in responses to faces.

tually requires a fundamental rethinking of how to ask questions and generate hypotheses. Within the traditional univariate framework, a researcher might ask a question like: "which parts of the brain are more active for perceiving humans compared to animals?" On the other hand, within an MVPA framework, a question would instead be: "where does information in the brain dissociate perceiving humans from animals?" Notice that the MVPA framework does not make assumptions about the direction (whether the brain area showed more or less activation to resting brain activity) nor the strength of the responses, but rather allows the patterns of activation to inform the representation of the two perceptions. In other words, univariate methods ask about the location of the activation and multivariate methods ask about the patterns of activation. Because the mean signal is removed in multivariate methods, these methods provide independent information about the nature of brain activation. As such, MVPA methods provide answers to questions about not only the parts of the brain involved with a given mental process, but also the implementation of mental processes within them. The relatively greater level of detail and sensitivity about the function and structure of the human brain afforded by MVPA can be more useful than univariate methods to inform theory in social and personality psychology.

One of the main challenges in cognitive neuroscience is the problem of reverse inference (Poldrack, 2011). Many psychologists using neuroimaging not only want to show which parts of the brain are involved with a mental process ("forward inference"), but also want to be able to infer "backwards" that, when a pattern of activation is apparent in the data, a given mental process is being engaged. Because there is no simple one-to-one mapping of activation within a region to complex social cognitive processes, the mere observation of activation in a particular region to a given stimulus is not sufficient to infer that a particular mental process is being engaged. However, because MVPA quantifies how strongly a pattern of activity correlates with a mental process, it provides a framework for developing a formal means to implement reverse inference (Poldrack, 2011). As such, MVPA methods provide a tractable approach for testing the neural mechanisms of how psychological processes are represented in the brain in a way that is better aligned with our understanding of the brain as a massively parallel processing system.

Because of the possibility for more accurate reverse inference, MVPA has greater promise for informing social and personality psychological theorizing than univariate approaches. However, effectively applying these methods also requires a greater degree of sophistication in the understanding of both quantitative methods and neurobiological systems. In the next section, we describe different ways to implement MVPA. Each of these approaches is powerful, but they can only be as useful as the psychological theories and paradigms on which they are based. A deep understanding of psychological and analytical sides is required to make fruitful connections between the levels of analysis. Thoughtfully and intentionally bringing multivariate approaches into social and personality psychology research will not only help advance psychological theories, but in turn help neuroscientists better understand the organization of the brain based on our best understanding of social cognition, personality, and behavior.

Varieties of Multivariate Pattern Analysis

MVPA is an umbrella term used to capture a variety of methods that use brain responses that are distributed across multiple voxels to make inferences about mental representations and other psychological processes. The most common MVPA methods used in social neuroscience have a variety of applications, though they all utilize multivariate fMRI activity measured across multiple voxels.

Classification.

In a typical fMRI experiment, the stimulus categories are treated as the independent variables to make inferences about brain responses, the dependent variable. MVPA classification turns this equation around. Instead of using stimulus categories to predict brain responses, classification approaches use the multi-voxel brain response patterns as the independent variable to make predictions about the stimulus categories (i.e., classify them). This was the approach used in the paper described above where Haxby and colleagues (2001) determined if activation patterns in the FFA could correctly dissociate multiple visual stimulus categories besides just faces.

MVPA classification studies aim to test whether patterns of brain activation can reliably classify, or dissociate, two or more mental states or stimulus types (Wagner et al., 2019). Contemporary MVPA classification studies now use machine learning algorithms, though early classification work used familiar statistical methods such as correlation analysis, linear discriminant analysis, or logistic regression. The most commonly used machine learning methods in MVPA classification studies include support vector machines, naive Bayes, and random forest models. There are a variety of differences in these algorithms, the details of which are beyond the scope of this chapter. However, each of these and other similar algorithms were imported from statistics and computer science fields and the choice of which to use can depend on a variety of factors related to the type of data being collected and the availability of computational resources for processing those data. In practice, however, there usually are not large differences in the conclusions drawn from using any of the standard machine learning algorithms and most have been demonstrated to have greater sensitivity for dissociating among category types (Haxby, 2012).

As with the application of machine learning methods in any area, one concern is that the increased sensitivity to the underlying signal also increases the vulnerability of overfitting responses to noise. If a machine learning algorithm detects a pattern in the data, it will leverage that pattern as part of its prediction even if the pattern is unrelated to the true signal. For this reason, common practice is to make inferences on the basis of an out-of-sample prediction model, where classification algorithms are first trained on subsets of the data (e.g., 75% of it) before being evaluated on independent portions of the data that are left out of the original analyses (e.g., the remaining 25%). For example, suppose a researcher is interested in using brain responses to classify which of four different emotional facial expressions a participant is viewing (e.g., happy, angry, surprised, disgusted) collected in four separate runs of the paradigm. Typically, the classification algorithm would be trained to dissociate each of the categories in three runs of the fMRI paradigm before being tested on the held-out run for accuracy. This train-then-test procedure is then repeated iteratively with each run serving as the testing data set. The classification results are then aggregated across each iteration and reported as average accuracy scores for each participant with the standard deviation indicating variation in accuracy across runs. Finally, this is repeated for all participants and final statistical analyses are performed to determine if accuracy scores are consistent across participants, typically using a non-parametric t-test against chance-level performance. Non-parametric methods do not make the same assumptions of normality as standard parametric t-tests. Because classification accuracies are typically not normally distributed, non-parametric methods are important for robust inferences. Classification algorithms often leverage within-subjects designs to capitalize on the large quantity of data per subject and control for person-to-person variation, so they do not necessarily require greater sample sizes than univariate methods. However, they may require more data or runs per subject (Coutanche & Thompson-Schill, 2012).

MVPA classification studies have been employed in a variety of different paradigms to study social cognition. In one study by Hassabis et al., (2014) participants were instructed to learn the personalities of four novel individuals and imagine how each would behave in different scenarios. Using MVPA classification, the authors were able to accurately identify which person was being imagined based on activity patterns in a portion of the medial prefrontal cortex. Although univariate methods had suggested this portion of the brain was associated with social cognition in general (Denny et al., 2012), the Hassabis study demonstrated that multivariate patterns of activity can not only dissociate thinking about people versus not thinking about people, but also identify information about specific individuals.

Beyond identifying regions of the brain that contain detailed information about social cognitive processing, classification analyses can also be used to test psychological theory. For example, people often describe their personal relationships in terms that are literally about physical space: "I am very close with my sister" or "my best friend and I have drifted apart lately." Is this just a metaphorical quirk of the way we use language, or do our minds use the same mechanisms to compute distances across a variety of domains? Parkinson et al. (2014) answered these questions in a study where participants viewed stimuli that varied in spatial distance (close vs. far), temporal distance (sooner vs. later), and social distance (friend vs acquaintance). In a clever use of MVPA classification, these researchers trained on one domain and tested in the other two for each iteration of this design. For example, a classifier was trained to differentiate sooner from later and then tested on stimuli that varied on social distance. If there is a reliable decision boundary between closer and further distances across conceptual domains then the cross-domain MVPA classification approach will identify parts of the brain that are shared across spatial, temporal, and social domains. Indeed, Parkinson et al., (2014) found that the right inferior parietal lobule – an area thought to be involved in sensorimotor transformations of information - showed consistent cross-domain classification across all three domains. Though previous studies had shown that this region was important for computing information related to numeric processing (Eger et al., 2009), univariate methods limited the ability to test hypotheses about how these this processing might be serving social cognitive processing too. This finding provides evidence that there is a common cortical mechanism supporting the dissociation between each of the three domains and suggests that the construal of interpersonal distance is built from more general cognitive mechanisms serving multiple psychological processes.

To summarize, MVPA classification studies provide a robust method for decoding categories of experimental conditions. Although it often requires more data for cross-validation, MVPA classification is more sensitive than traditional univariate approaches in decoding information about mental states and can reveal information that is represented in distributed patterns of responses rather than simple magnitude differences.

Representational Similarity Analysis.
Classification methods in MVPA are powerful techniques for decoding category-level information with a given brain region. However, these techniques are not optimal for directly comparing hypothesized models (Popal et al., 2019). For example, it may be possible to use MVPA to accurately classify human faces, animal faces, and inanimate objects. However, if your theories suggests that human and animal faces should be more related to each other than they are to the inanimate objects, then classification accuracies alone cannot address this question. Another MVPA technique called representational similarity analysis (RSA) addresses some of the limitations of MVPA classification and is rapidly gaining popularity in MVPA studies in social neuroscience.

RSA is a procedure for measuring the conceptual distances among categories by quantifying the similarities of their voxel-wise activity patterns (Kriegeskorte, Mur, & Bandettini, 2008). By abstracting away from the fMRI signal into similarity space (i.e., distance or correlation matrices of voxel-by-voxel responses), researchers can quantitatively describe the relationships among each of the categories and test hypotheses about similarity that are derived from any other kind of information, including behavioral data or theoretical models. For example, suppose a researcher is interested in whether participants perceive animal faces as more human-like than inanimate object categories. Rather than just asking if an algorithm can accurately classify each condition (as in MVPA classification), the researcher can calculate the correlation distances (e.g., 1 - Spearman r) among patterns of responses to each stimulus within a given region of the brain to create a matrix that indicates the magnitude of the difference in neural response between the categories, known as a neural dissimilarity matrix (Figure 2a). Once these neural dissimilarity matrices are calculated, dissimilarity values between pairs of stimuli can then be related against any number of predicted similarity structures derived from competing theoretical models (Figure 2b) or separate behavioral data. The models can then be formally compared in a regression framework to determine which one most accurately describes the data.

At its core, RSA is a straightforward method: calculate the similarity between brain responses to stimuli and relate those to similarity structures derived from other sources. The power of RSA derives from its ability to abstract fMRI signals into similarity space among stimuli, which affords a comparison across any modality that can produce a similarity matrix among the same stimuli. For example, in an early paper using RSA, Kriegeskorte, Mur, Ruff, et al. (2008) investigated whether brain responses to different object categories were shared between human fMRI data and single-cell electrical recordings in monkeys. For example, do humans and monkeys make similar differentiations between nature and artificial objects, and animate and inanimate objects. Because the humans and monkeys viewed all the same objects, a common similarity structure could be constructed within each species using their respective brain measure modality. Indeed, despite differences in brain measure modality, the authors were able to relate these two sources of data to find a strong correspondence between the two species. These results demonstrate that there may be relatively preserved object categories that are shared across species and not modified by language or other species-specific cognitive faculties. This study underscores how useful and flexible leveraging similarity spaces can be for describing the organization of mental constructs in the brain by allowing for fusing of data that would otherwise be impossible to compare.

Of course, RSA also brings unique challenges and decision points. One issue is the level of granularity at which to calculate a neural dissimilarity matrix. RSA methods make it possible to calculate similarity matrices at the item-level (Mumford et al., 2014) as well as calculate dissimilarity matrices at the category-level by aggregating across multiple trials within categories. The decision to choose item-level or category-level similarity depends on a number of factors related to the design of the study, the amount of data collected, and the details of the hypothesis being tested. Similar to behavioral paradigms, the more trials per condition the greater power there will be to estimate responses to that condition and be able to generalize to other exemplars from those conditions. Another issue is the scope of the brain areas you want included in the calculation of the neural similarity matrices. An RSA framework allows dissimilarity measures to be calculated at any scale: small regions, large regions, and even non-adjacent areas analyzed together (Jolly & Chang, 2021). Including too many areas that are not involved in processing the stimuli at hand will distort the similarity matrices

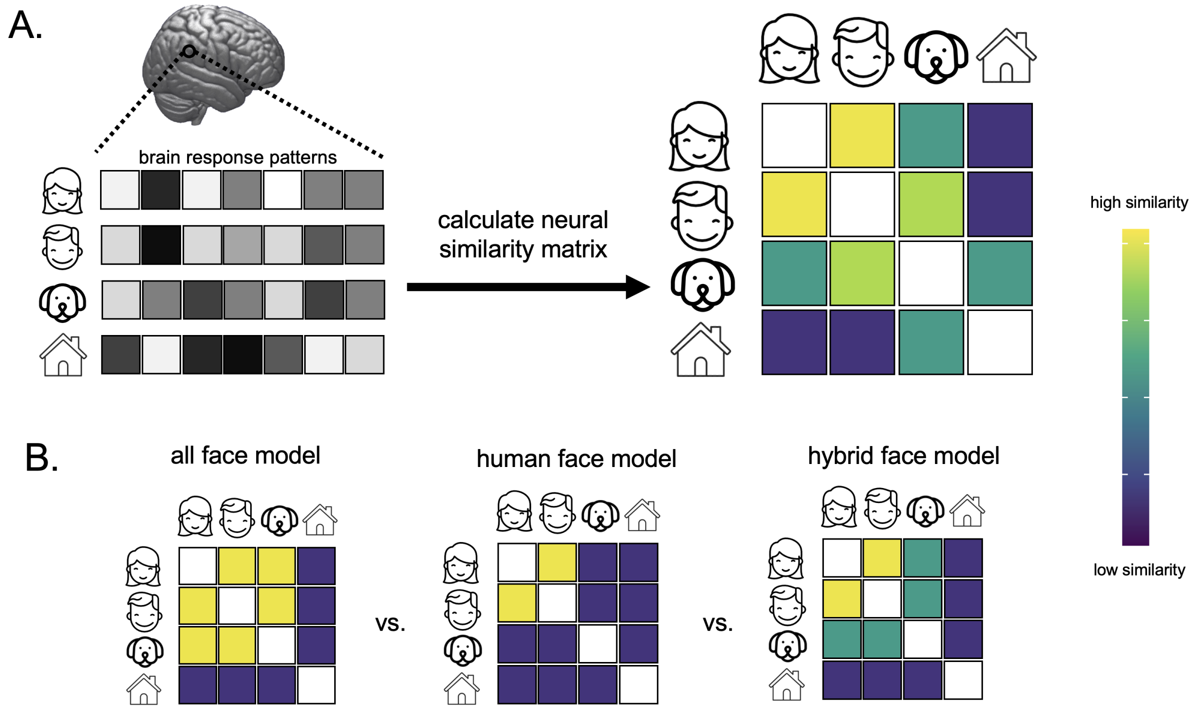

Fig. 2. *General Approach for Representational Similarity Analysis.* An example of a general approach for conducting a representational similarity analysis (RSA). (a) Voxelwise brain responses for each stimulus are correlated with one another to compute a pairwise neural similarity matrix. (b) Researchers can then generate any number of theoretically informed or empirically measured similarity models to compete in a regression framework for which best explains the similarity in neural responses.

by including areas that are not contributing to the processing of the mental phenomenon being studied. On the other hand, narrowing down too much may not accurately capture how distributed brain regions work together to generate mental representations of more complex mental processes typically studied by social and personality psychologists. As such, to properly utilize the advantages of RSA, it is critical to have a theoretically informed understanding of which brain regions are likely to be involved in the processes being studied. Finally, there are a number of statistical issues to be addressed when using similarity matrices as the basis for modeling, including non-independent and nested data structures, collinearity of competing models, and the appropriate threshold for the significance of a correlation within a matrix.

Despite the challenges that come with using RSA, several studies have successfully utilized these methods for generating novel insights on a variety of phenomena of interest to social and personality psychologists. One study by Thornton & Mitchell (2017) used RSA to demonstrate that specific, familiar individuals elicited consistent patterns of similarity across participants in areas of the brain that have previously been shown to be important for social cognition. Moreover, the researchers also found that neural similarity patterns dovetailed with similarity patterns from theories of social cognition (e.g., stereotype content model, Big 5 personality traits). In another study, Stolier & Freeman, (2016) found that the participants' biases in the perception of social categories (i.e., race, gender, and emotional expression) were reflected in the similarity structure of the fusiform gyrus and orbitofrontal cortex which are typically involved in basic processing of visual facial stimuli. Using the model comparison approach, these researchers were also able to show that their effects held even after accounting for the lower-level visual similarities between facial stimuli. These results suggest that stereotypes and similar social-conceptual biases influence perceptual processing at lower levels than previously identified.

The RSA method has great potential for building a more comprehensive understanding of how the brain represents social information based on both robust theoretical and data-driven models. Yet, reframing hypotheses in terms of the similarity among conditions requires a pivot away from traditional modes of hypothesis generation and data analysis.

Biomarker & Neural Signature Analyses.

Among the most robust findings across years of neuroscience research is the lack of functional specificity of individual brain regions. In the early days of social neuroscience, researchers looked to find regions that would correspond one to one with a psychological process – for example, finding the "fear" region or the "self" region of the brain. Yet, it has become increasingly clear that most mental processes – especially the ones that social psychologists are interested in - depend on activity in numerous brain regions has made it difficult to identify mental processes associated with particular brain areas or ascribe multifaceted mental processes to them. MVPA methods such as the ones described above more realistically reflect that neural representations of certain mental processes are better captured by broader and more distributed brain systems than those identified using univariate methods. Another type of MVPA approach extends the scope of these distributed neural representations to include the entire brain at once. These approaches—referred to as "biomarker" or neural signature analyses—capitalize on these broadly distributed activation patterns to generate whole-brain predictive models of mental state representations, clinical diagnoses, or individual differences. When executed carefully and validated appropriately, these models are powerful tools because they provide quantitative, falsifiable predictions and tend to generate greater effect sizes and specificity than other methods (Kragel et al., 2018).

Biomarker-based models leverage the idea that complex behavioral phenotypes will be supported by regions distributed throughout the brain. Like MVPA classification analyses, whole-brain biomarkers are typically built using a variety of modern machine learning techniques that use voxelwise brain activity as the predictors for a behavioral outcome (Woo et al., 2017). However, unlike typical MVPA classification analyses that are cross-validated across runs, whole-brain biomarker models are trained within a subset of individuals and generate a predictive model that can be applied to completely independent groups of participants. This allows biomarker models to be tested across a wider range of people than the sample on which it was trained. As such, biomarker studies allow for generalizability across different participant demographics, different scanners, and even different experimental paradigms. An additional benefit is related to interpretability: once a biomarker or signature is trained, it is also possible to inspect the model to see what parts of the brain contribute most strongly to the predictive accuracy of the model and the direction of their relationships to the outcome.

Biomarker models can be used to study a wide variety of psychological phenomena of interest. One early biomarker study was conducted by Wager et al. (2013) to investigate the distributed neural representations of physical pain. The subjective experience of pain to a noxious stimulus is difficult to assess with self-reports or other physiological indicators (Davis et al., 2020). In this study, Wager et al., (2013) trained a whole-brain predictive model that was able to accurately predict pain levels in an independent sample of participants. Specifically, the researchers used a regression-based machine learning algorithm based on every voxel in the brain to identify patterns of fMRI responses that were associated with levels of heat-induced pain. First, they built the whole-brain signature by fitting the model to a group of participants experiencing four levels of pain intensity. They showed that the signature showed very high sensitivity and specificity (greater than 94%) in predicting pain intensity level. Next, they cross-validated these results in an independent sample of participants and again found highly accurate predictions of pain intensity. Finally, in two more independent samples of participants, they further specified the robustness of the neurological pain signature by showing that it was specific to the increases in experiences of pain intensity and not the anticipation of pain, the recall of previous pain, or other mental states that are known to activate similar brain networks (e.g., social rejection). This study demonstrates the utility of using a neural signature approach but also underscores the importance of generalizing the findings across independent participants and validating the measure against similar constructs.

Another study illustrates how the biomarker approach can be useful to social and personality psychology. Many studies in social psychology use static emotionally valenced images to elicit the perception of negative affect in participants. However, researchers typically rely on self-report to assess the emotional intensity of the images. It would be useful to develop an alternative method that less intrusively or implicitly assesses the level of emotional intensity evoked by a stimulus that generalizes across participants. To this end, Chang et al., (2015) set out to test a whole-brain biomarker for predicting what they called a picture-induced negative affect signature (PINES). Using a procedure that mirrored the approach used in the neurological pain signature from Wager et al., (2013), Chang et al., (2015) found that the PINES model was sensitive at predicting the level of negative affect evoked by a given image. The PINES model was specific to the perception of images: it did not predict the affective responses to pain or vice versa. Moreover, a central message of the study was the observation that these results could not be reduced to any individual brain structure within the regions identified in the signature, underscoring the importance of using a multivariate approach in this case.

Another advantage of using biomarker models is the ease by which they can be readily shared among researchers, tested independently, and potentially integrated into a broader predictive framework. The signatures that are derived from these procedures produce whole-brain images of statistical parameters – called voxelwise weight maps – that can be saved in standard neuroimaging formats and applied to any data that are aligned to the same coordinate space to produce a prediction to any set of data based on the signature. This provides a potentially powerful framework to combine signatures derived from different psychological processes or constructs into integrative predictive models at multiple levels of abstraction (Woo et al., 2017).

Towards the Future

Integrating Multivariate Neuroimaging with other Methods and Designs.

The MVPA methods described above provide a suite of approaches and tools that leverage distributed brain activation patterns to understand how neural systems give rise to complex social and personality processes. Each of these tools is powerful in its own right and provides new avenues of understanding the link between social behavioral and brain mechanisms. MVPA methods need not stand on their own, however, and can be incorporated within other sets of methods and rich study designs.

Social network analysis is becoming a popular tool for quantifying the structure of social relationships within a variety of contexts. In a pioneering study combining social network analysis with MVPA methods, Parkinson et al., (2017) recruited an entire incoming academic cohort—275 first-year Masters of Business Administration students—and had them identify other members within their cohort with whom they socialized or knew. This procedure yielded a map of the social relationship structure of the entire network that could be used to calculate social network analysis metrics. The authors then recruited a subset from this cohort for an fMRI scanning session where participants viewed short videos of their classmates describing themselves to estimate brain responses to group members. Finally, social network analysis metrics were converted into dissimilarity values and related to the similarity of multivariate fMRI responses using RSA. The researchers found that each of the social network metrics were spontaneously encoded within these participants but in different regions of the brain, indicating that people track social networks in a complex, abstract way that would be difficult to detect using univariate methods.

Studies of interpersonal perception have benefited from the rich nature of so-called "round-robin" designs (Kenny & Albright, 1987). In these designs, every participant is both a perceiver and social target for every other member of their group, and participants are asked to make trait judgments of members of their social groups. This allows researchers to measure a variety of interpersonal perception characteristics such as interpersonal consensus, self-other agreement, and dyadic similarity. These designs can also be used in a neuroimaging context. Guthrie et al. (2022) ran a round-robin fMRI study investigating whether the social relationship strength between pairs of individuals could be predicted by their shared brain responses when thinking of other members of their group. To test this, the researchers recruited twenty groups of six participants each to undergo scanning while thinking of each other member of their group. Next, the investigators used multivariate similarity responses to measure the agreement in brain-to-brain responses between pairs of individuals when they are thinking of the other people in their group. They found that similarity in the brain-to-brain responses within regions previously implicated in mentalizing and social motivation (e.g., the dorsal medial prefrontal cortex and anterior insula) consistently predicted social relationship strength between pairs of individuals within each group. These findings suggest that the more similarly two people's brains process socially relevant information about their group member, the more likely those individuals are to have strong interpersonal relationships.

The studies above are just a couple of examples of the ways in which MVPA methods can be integrated with innovative study designs to provide insights and analytic opportunities that would not otherwise be possible. Though these kinds of designs can be technically and logistically challenging to conduct, they provide tremendous opportunities for the advancement of our understanding of the social brain.

Beyond fMRI.

We focused the discussion in this chapter of MVPA methods to how they have been used in fMRI. However, multivariate methods can be applied across any neuroimaging modality in which there are multiple signals that can be incorporated together to test hypotheses or compare models.

There are now several papers applying MVPA methods to the study of social behavior and cognition using electroencephalography (EEG). In one study, Hundrieser et al., (2021) used an MVPA classification approach to predict participants' binary yes/no agreement with the moral acceptability of various statements (e.g., "wars are acceptable") based on spatiotemporal patterns from 61 electrode channels. Capitalizing on EEGs superior temporal resolution compared to fMRI, these researchers were able to determine not only that the EEG responses could classify the behavioral response but also identify the epochs in the time series when accurate classification was possible.

Another promising direction using multivariate EEG is the development of microstate event-related potentials (ERPs). ERPs use EEG signals averaged across many trials to estimate the waveform response to a given stimulus. Microstate ERP analysis is an approach for identifying stable configurations of global electric brain activity ERP using information from all of EEG electrodes in a multivariate framework (Cacioppo et al., 2014). This method has the advantage over standard ERP approaches of being able to identify microstates of temporal stability in a data-driven fashion over multiple electrodes to capture reliable effects missed with standard EEG/ERP approaches. In other words, rather than treating each millisecond of electrical signal from each recording electrode separately, this analysis looks for stable patterns of electrical activity across the electrodes that remain stable for a period of time. This approach has been used to successfully capture both the similarities and differences in brain states short-term vs. long-term romantic intentions following the viewing of photographs of attractive strangers (Cacioppo, Bolomont, & Monteleonem, 2017).

Relative to fMRI, there has been less widespread use of multivariate EEG/ERP methods within social neuroscience. However, testing hypotheses related to the timing of psychological processes engaged during social cognition would be difficult, if not impossible, to test with the slow temporal resolution of fMRI. EEG measures are much better suited to questions about temporal information and applying MVPA methods to these questions is potentially a very fruitful opportunity to make deep insights into these domains.

Although fMRI is the most common MRI modality used in social neuroscience, there are a variety of other MRI-based modalities that measure different properties of the brain, for example structural differences in grey or white matter. Voxel-Based Morphometry (VMB) methods allow researchers to examine structural differences in tissue concentration that is typically used for examining voxel-wise grey matter concentrations. Diffusion MRI (dMRI) is a structural neuroimaging technique that measures the anatomical (white matter) connectivity between regions. The relatively static nature of dMRI measures make them particularly well suited for studying stable behavioral characteristics, such as personality traits. Because the breadth of behaviors and characteristics underlying many personality dispositions is very broad, it is likely that these processes cannot be localizable to any single brain region but rather are distributed across various brain systems. Variability in the communication between these systems may explain some of the mechanisms underlying individual differences in personality characteristics. Utilizing biomarker modeling methods, it is possible to build predictive models of individual differences in personality based on distributed anatomically connectivity metrics. Chavez et al. (2022) built a biomarker model of individual differences in self-esteem based on the results of dMRI from a previous study. The researchers then applied the predictions from this model to an independent sample of participants from a completely new study to generalize these results across time, scanners, and subject demographics. They found that a multivariate model of self-esteem outperformed univariate models within the same predictive modeling framework. More broadly, similar predictive multivariate dMRI effects can also be seen in models of the Big Five personality traits that generalize across samples of participants (Stendel & Chavez, in press).

Despite the fact that MVPA methods were pioneered in the context of fMRI data, we anticipate that researchers will recognize that these methods are built on more general analytic frameworks and apply them across any number of contexts where multivariate information is assessed. Moreover, the potential for shared methodological approaches across levels of analysis—from brain to cognition to behavior—underscores the utility of the multivariate approach in social neuroscience.

Conclusion

Social neuroscience straddles lines between social and personality psychology and cognitive neuroscience with the promise of mutual informativeness. Just as theories and methods in social psychology have increased in their detail, nuance, and sophistication, so too has our understanding of the brain and the methods we use to examine it. Social psychology has long emphasized context and situations as a critical way of understanding social behavior. Multivariate methods mirror this at the level of the brain. It is not enough to just know if a region is involved in a process; it is necessary to know how other regions are active simultaneously to embed the activity of a given area within its neural context (McIntosh, 1998).

In this chapter, we have provided an overview of various approaches to multivariate neuroimaging (i.e., MVPA methods) for social and personality psychologists. Several high-quality reviews have also been published that cover in greater detail each of these methods and how they may be applied to topics of interest to social psychologists (see: Popal et al., 2019; Wagner et al., 2019; Weaverdyck et al., 2020). However, each of the methods covered here are being actively developed and optimized for best practices. We suspect that optimal methods for studying one topic (e.g., lower-level perception) may not always transfer to other topics (e.g. higher-order social cognition). However, this underscores the necessity for social and personality psychologists to be involved with these efforts from the outset. The fundamental mechanism of our social behavior lies in the three-pound organ encased within the skull. However, the details of how this system gives rise to all the complex social and personality processes in which we are interested remains a complex challenge and entangled mystery (Chavez, 2021). Multivariate neuroimaging methods provide a powerful path forward and will likely become the dominant approach to understanding the brain basis of social cognition and behavior for years to come.

References

Berkman, E. T., Cunningham, W. A., & Lieberman, M. D. (2014). Research Methods in Social and Affective Neuroscience. In H. T. Reis & C. M. Judd (Eds.), Handbook of Research Methods in Social and Personality Psychology (2nd ed., pp. 123–158). Cambridge University Press. https://doi.org/10.1017/CBO9780511996481.011

Cacioppo, J. T., & Berntson, G. G. (1992). Social psychological contributions to the decade of the brain. Doctrine of multilevel analysis. The American Psychologist, 47(8), 1019–1028.

Cacioppo, J. T., Berntson, G. G., Lorig, T. S., Norris, C. J., Rickett, E., & Nusbaum, H. (2003). Just because you're imaging the brain doesn't mean you can stop using your head: A primer and set of first principles. Journal of Personality and Social Psychology, 85(4), 650–661. https://doi.org/10.1037/0022-3514.85.4.650

Cacioppo, S., Weiss, R. M., Runesha, H. B., & Cacioppo, J. T. (2014). Dynamic spatiotemporal brain analyses using high performance electrical neuroimaging: theoretical framework and validation. Journal of Neuroscience Methods, 238, 11-34. https://doi.org/10.1016/j.jneumeth.2014.09.009

Cacioppo, S., Bolmont, M., & Monteleone, G. (2018). Spatio-temporal dynamics of the mirror neuron system during social intentions. Social Neuroscience, 13(6), 718-738. https://doi.org/10.1080/17470919.2017.1394911

Chang, L. J., Gianaros, P. J., Manuck, S. B., Krishnan, A., & Wager, T. D. (2015). A Sensitive and Specific Neural Signature for Picture-Induced Negative Affect. PLOS Biology, 13(6), e1002180. https://doi.org/10.1371/journal.pbio.1002180

Chavez, R. S. (2021). Tangled Representations of Self and Others in the Medial Prefrontal Cortex. In M. Gilead & K. N. Ochsner (Eds.), The Neural Basis of Mentalizing (pp. 599–611). Springer International Publishing. https://doi.org/10.1007/978-3-030-51890-5$_3$1

Chavez, R. S., Tovar, D. T., Stendel, M. S., & Guthrie, T. D. (2022). Generalizing effects of frontostriatal structural connectivity on self-esteem using predictive modeling. Cortex, 146, 66–73. https://doi.org/10.1016/j.cortex.2021.09.021

Coutanche, M. N., & Thompson-Schill, S. L. (2012). The advantage of brief fMRI acquisition runs for multi-voxel pattern detection across runs. Neuroimage, 61(4), 1113-1119.

Davis, K. D., Aghaeepour, N., Ahn, A. H., Angst, M. S., Borsook, D., Brenton, A., Burczynski, M. E., Crean, C., Edwards, R., Gaudilliere, B., Hergenroeder, G. W., Iadarola, M. J., Iyengar, S., Jiang, Y., Kong, J.-T., Mackey, S., Saab, C. Y., Sang, C. N., Scholz, J., … Pelleymounter, M. A. (2020). Discovery and validation of biomarkers to aid the development of safe and effective pain therapeutics: Challenges and opportunities. Nature Reviews Neurology, 16(7), 381–400. https://doi.org/10.1038/s41582-020-0362-2

Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. Journal of Cognitive Neuroscience, 24(8), 1742–1752. https://doi.org/10.1162/jocn$_{a0}$0233

Duchaine, B., & Yovel, G. (2015). A Revised Neural Framework for Face Processing. Annual Review of Vision Science, 1(1), 393–416. https://doi.org/10.1146/annurev-vision-082114-035518

Eger, E., Michel, V., Thirion, B., Amadon, A., Dehaene, S., & Kleinschmidt, A. (2009). Deciphering cortical number coding from human brain activity patterns. Current Biology, 19(19), 1608-1615.

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. Science, 302(5643), 290-292.

Guthrie, T. D., Benadjaoud, Y. Y., & Chavez, R. S. (2022). Social Relationship Strength Modulates the Similarity of Brain-to-Brain Representations of Group Members. Cerebral Cortex, 32(11), 2469–2477. https://doi.org/10.1093/cercor/bhab355

Hassabis, D., Spreng, R. N., Rusu, A. a, Robbins, C. a, Mar, R. a, & Schacter, D. L. (2014). Imagine All the People: How the Brain Creates and Uses Personality Models to Predict Behavior. Cerebral Cortex, 24(8), 1979–1987. https://doi.org/10.1093/cercor/bht042

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(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: the early beginnings. Neuroimage, 62(2), 852-855. https://doi.org/10.1016/j.neuroimage.2012.03.016

Hundrieser, M., Mattes, A., & Stahl, J. (2021). Predicting participants' attitudes from patterns of event-related potentials during the reading of morally relevant statements – An MVPA investigation. Neuropsychologia, 153, 107768. https://doi.org/10.1016/j.neuropsychologia.2021.107768

Jimura, K., & Poldrack, R. A. (2012). Analyses of regional-average activation and multivoxel pattern information tell complementary stories. Neuropsychologia, 50(4), 544-552.

Jolly, E., & Chang, L. J. (2021). Multivariate spatial feature selection in fMRI. Social Cognitive and Affective Neuroscience, 16(8), 795–806. https://doi.org/10.1093/scan/nsab010

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. The Journal of Neuroscience, 17(11), 4302–4311.

Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. Psychological Bulletin, 102(3), 390–402. https://doi.org/10.1037/0033-2909.102.3.390

Kragel, P. A., Koban, L., Barrett, L. F., & Wager, T. D. (2018). Representation, Pattern Information, and Brain Signatures: From Neurons to Neuroimaging. Neuron, 99(2), 257–273. https://doi.org/10.1016/j.neuron.2018.06.009

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. Frontiers in Systems Neuroscience, 2, 4. https://doi.org/10.3389/neuro.06.004.2008

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. Neuron, 60(6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043

McIntosh, A. R., Bookstein, F. L., Haxby, J. V., & Grady, C. L. (1996). Spatial Pattern Analysis of Functional Brain Images Using Partial Least Squares. NeuroImage, 3(3), 143–157. https://doi.org/10.1006/nimg.1996.0016

McIntosh, A. R. (1998). Understanding neural interactions in learning and memory using functional neuroimaging. Annals of the New York Academy of Sciences, 855(1), 556-571. https://doi.org/10.1111/j.1749-6632.1998.tb10625.x

Mumford, J. A., Davis, T., & Poldrack, R. A. (2014). The impact of study design on pattern estimation for single-trial multivariate pattern analysis. NeuroImage, 103, 130–138. https://doi.org/10.1016/j.neuroimage.2014.09.026

Parkinson, C., Kleinbaum, A. M., & Wheatley, T. (2017). Spontaneous neural encoding of social network position. Nature Human Behaviour, 1(5), 0072. https://doi.org/10.1038/s41562-017-0072

Parkinson, C., Liu, S., & Wheatley, T. (2014). A common cortical metric for spatial, temporal, and social distance. The Journal of Neuroscience, 34(5), 1979–1987. https://doi.org/10.1523/JNEUROSCI.2159-13.2014

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. Neuron, 72(5), 692–697.

Popal, H., Wang, Y., & Olson, I. R. (2019). A Guide to Representational Similarity Analysis for Social Neuroscience. Social Cognitive and Affective Neuroscience, 14(11), 1243–1253. https://doi.org/10.1093/scan/nsz099

Stendel, M., & Chavez, R. S. (in press). Beyond the brain localization of complex traits: Distributed white matter markers of personality. Journal of Personality.

https://doi.org/10.1111/jopy.12788

Stolier, R. M., & Freeman, J. B. (2016). Neural pattern similarity reveals the inherent intersection of social categories. Nature Neuroscience, 19(6), 795–797. https://doi.org/10.1038/nn.4296

Thornton, M. A., & Mitchell, J. P. (2017). Consistent Neural Activity Patterns Represent Personally Familiar People. Journal of Cognitive Neuroscience, 29(9), 1583–1594. https://doi.org/10.1162/jocn$_{a0}1151$

Wager, T. D., Atlas, L. Y., Lindquist, M. A., Roy, M., Woo, C.-W., & Kross, E. (2013). An fMRI-Based Neurologic Signature of Physical Pain. New England Journal of Medicine, 368(15), 1388–1397. https://doi.org/10.1056/NEJMoa1204471

Wagner, D. D., Chavez, R. S., & Broom, T. W. (2019). Decoding the neural representation of self and person knowledge with multivariate pattern analysis and data-driven approaches. WIREs Cognitive Science, 10(1), e1482. https://doi.org/10.1002/wcs.1482

Weaverdyck, M. E., Lieberman, M. D., & Parkinson, C. (2020). Multivoxel pattern analysis in fMRI: A practical introduction for social and affective neuroscientists. Social Cognitive and Affective Neuroscience, 15(4), 487–509. https://doi.org/10.1093/scan/nsaa057

Woo, C.-W., Chang, L. J., Lindquist, M. A., & Wager, T. D. (2017). Building better biomarkers: Brain models in translational neuroimaging. Nature Neuroscience, 20(3), 365–377. https://doi.org/10.1038/nn.4478

Woo, C. W., Koban, L., Kross, E., Lindquist, M. A., Banich, M. T., Ruzic, L., ... & Wager, T. D. (2014). Separate neural representations for physical pain and social rejection. Nature Communications, 5(1), 5380.