RIVISTA INTERNAZIONALE DI FILOSOFIA E PSICOLOGIA
DOI: 10.4453/rifp.2022.0018
ISSN 2039-4667; E-ISSN 2239-2629
Vol. 13 (2022), n. 3, pp. 187-199

RICERCHE

Representations and processes: What role for multivariate methods in cognitive neuroscience?

Davide Coraci[(α), (β)]

Ricevuto: 31 gennaio 2022; accettato: 12 dicembre 2022

Abstract The significance of neuroscientific findings for the analysis of central problems in cognitive science has long been a matter of debate. Recent developments in cognitive neuroscience have reignited this discussion, especially with regard to the study of cognitive representations and cognitive processes. The present paper focuses on multivariate analyses, a class of neuroscientific methods that promises to shed new light on the neural bases of cognitive representations. Multivariate approaches are both powerful and increasingly used. Yet, we argue that their successful application in neuroscience requires significant theoretical and methodological clarification. After providing a preliminary assessment of the pros and cons of multivariate methods, we claim that their successful application crucially depends on how we conceptualize the relationships between representations, cognitive processes, and neural data, in other words, on the cognitive ontology we use to describe the human mind. Our discussion also highlights some general strengths and weaknesses of neuroscientific contributions to the program of classical cognitive science.
KEYWORDS: Cognitive Process; fMRI; Multivariate Analysis; Marr's Three Levels of Analysis; Cognitive Ontology

Riassunto *Rappresentazioni e processi: quale ruolo per i metodi multivariati nelle neuroscienze cognitive?* - La rilevanza dei risultati neuroscientifici per quanto riguarda i problemi centrali delle scienze cognitive è motivo di discussione. Recenti sviluppi nelle neuroscienze cognitive hanno rianimato tale dibattito, in particolare rispetto allo studio delle rappresentazioni e dei processi cognitivi. Il presente articolo si focalizza sull'analisi multivariata, un insieme di metodi neuroscientifici che si promettono di studiare le basi neurali delle rappresentazioni cognitive. Nonostante le potenzialità e l'uso pervasivo degli approcci multivariati, in questo lavoro sosteniamo che prima di poter valutare il loro effettivo contributo nello studio delle rappresentazioni cognitive sia necessaria una chiarificazione teorica e metodologica. Dopo una discussione preliminare dei vantaggi e degli svantaggi dei metodi multivariati, evidenziamo come una loro efficace applicazione dipenda in maniera sostanziale da come viene intesa la relazione tra rappresentazioni, processi cognitivi e dati neurali o, in altre parole, dall'ontologia cognitiva che impieghiamo per descrivere la mente umana. Il presente lavoro affronta inoltre i generali punti di forza e gli elementi critici rispetto al contributo della ricerca neuroscientifica nel programma delle scienze cognitive classiche.
PAROLE CHIAVE: Cognitive Process; fMRI; Multivariate Analysis; Marr's Three Levels of Analysis; Cognitive Ontology

[(α)] Dipartimento di Neuroscienza Cognitiva, Computazionale e Sociale - IMT – Scuola di Alti Studi, Piazza San Ponziano, 6 – 55100 Lucca (IT)
[(β)] MoMiLab Research Unit – IMT – Scuola di Alti Studi, Piazza San Ponziano, 6 – 55100 Lucca (IT)

E-mail: davide.coraci@imtlucca.it (✉)

IN THE SECOND HALF OF the last century, cognitive science emerged as a research program aimed at unifying the study of the mind across various disciplines, integrating computational methods from computer science and neuroscience, as well as more traditional methods from psychology, philosophy, linguistics, and anthropology. In recent years, critics have raised concerns about various aspects of cognitive science. Some scholars have questioned the success of the enterprise and raised doubts about its viability and coherence as an interdisciplinary research program.[1] Others have instead criticized the role that some of the disciplines constituting cognitive science play within the broader research program. Yet others doubt whether neuroscience has made or can make a valuable contribution or, more specifically, if it can advance cognitive science, especially in regard to classical problems related to how the mind represents and processes information.[2]

In this connection, one major challenge concerns the concept of representation, that is, how the human mind processes and represents the external world. Neuroscience offers important opportunities for studying the neural substrates of representations. Explaining how "neural representations"[3] are related to, and influenced by, cognitive representations and processes would represent a key step in bridging the gap between neuroscience and theories in cognitive psychology, leading to a mature cognitive neuroscientific investigation of the human mind. However, the impact of neuroscientific findings on the study of representation is still an open question.

The present paper contributes to this ongoing debate by focusing on multivariate analyses, a specific class of recently developed neuroscientific methods used to analyze neural data. Multivariate methods differ from traditional methods in neuroscience and are considered a promising approach for shedding light on the concept of representation.[4] However, multivariate analyses raise methodological issues concerning the interpretation of their results. In particular, such issues concern the specific experimental features detectable by means of multivariate approaches. To better assess the relevance of multivariate methods for understanding how stimuli features are cognitively and neurally encoded by humans, a deeper analysis of the concept of representation is needed.

Our aim is to provide a preliminary assessment of multivariate methods as a tool for studying representations within cognitive (neuro)science, by highlighting the pros and cons of these techniques and emphasizing some of the methodological problems raised by their application. In a nutshell, we argue that multivariate methods are indeed a valuable tool for improving our understanding of cognitive representations and hence advancing the cognitive science program; however, their success crucially depends on how we conceptualize the relationships between cognitive states, cognitive processes, and neural data – a issue which is not still fully thematized and needs further work to be properly assessed. In this paper, we point to the main problems at the interface between neuroscientific research and classical reflections in cognitive science and offer a tentative assessment and discussion of their possible solutions.

The structure of the paper is as follows. We start with a quick recap of the notions of cognitive representation and process as discussed in cognitive science and neuroscience (Section 1). Then, in Section 2, we consider how multivariate analyses of neuroimaging data can shed light on the problem of representation. In Section 3, we use two toy-examples to illustrate several methodological issues that arise from the use of these approaches. As we shall see, the main problem is assessing whether neuroimaging results can be considered to provide evidence for representations, or instead address the cognitive processes that operate on them. As argued in Section 4, this issue is strictly related to the debate on the absence of a clear-cut "cognitive ontology" for describing human mental states.[5] We suggest that David Marr's "three levels of analysis"[6] model provides a useful conceptual framework to consider this central problem. In Section 5, we conclude with some remarks on the role of neuroscience within the cognitive science enterprise more generally.

1 Representations and processes in cognitive (neuro)science

Since the early days of cognitive science, the intertwined notions of cognitive representation and cognitive processing process have played key roles in the investigation of the mind. The main reason for this can be traced back to Computability Theory,[7] which gave cognitive science its central conceptual framework: the *Turing Machine* model, in which the (human) mind is conceptualised as a computational system in which cognitive processes are computations that operate on the internal symbols of the system.[8]

The pivotal role of internal symbols in performing cognitive tasks, as emphasized by Jerry A. Fodor in his famous slogan "no computation without representation",[9] has largely been discussed in the context of computational representational theories of mind. In these theories, the symbols upon which computations operate are representations, that is, internal (mental) entities with syntactic and semantic characteristics. Thus, representations are (mental) states characterized by a particular content that refers to the external world, making them semantically evaluable. Cognitive processes can then be understood as operations on representations.

The relationship between mental representation and computation, i.e., cognitive process, as discussed by Fodor finds clarification and systematization in Marr's model.[10] Marr proposes we conceptualize the information-processing performed by a system at three levels: as computation, algorithm, and implementation. This conceptualization additionally inspires helpful insights into the connections between these levels. The highest or computational level relates to the task performed by the system. The intermediate or algorithmic level concerns how computational theory is effectively carried out. Finally, the implementational level considers the physical substrates on which such information processing is realized.

Modern neuroscience addresses many of the traditional topics in cognitive science, employing technologies like electroencephalography (EEG), *Positron Emission Tomography* (PET), and *Magnetic Resonance Imaging* (MRI). MRI, the focus of this paper, allows neuroscientists to study the neural correlates of cognitive functions by tracing the electromagnetic behavior of oxygen in the blood flowing through brain regions. Indeed, when the MRI technique is used in an experimental setting, local consumption of oxygen is considered to provide a clue to the neural activity that results from the experimental manipulation. Here, we will talk about functional MRI or fMRI, which aims at detecting activity-related changes in the blood oxygenation of neural cells, that is, the "blood oxygenation level dependent" signal (BOLD).

The most common approach to assessing how the activity of brain regions is affected by experimental variables in fMRI-based studies is to detect variation in the BOLD signal at the level of single voxels[11] or voxel clusters. The peaks of the BOLD signal measured under different experimental conditions are first compared voxel by voxel and then aggregated into clusters that define brain regions. Analyses of neuroimaging data based on this approach are called "*univariate*" or "*voxel-based*" analyses. The univariate analysis contrasts BOLD signals (proxies for neural activity) across different experimental conditions, revealing whether a voxel or cluster of voxels was more or less active during one manipulation as compared to another. From this evidence, the neuroscientist can infer that the observed brain location was recruited for processing the experimental stimuli and, in turn, associate that locus with engagement of the supposedly recruited cognitive function.

However, as noted by Davis and Poldrack,[12] univariate approaches are poor tools for addressing the problem of representation. Since univariate analyses are based on the contrast of averaged brain activity associated with different experimental conditions (e.g., two different classes of stimuli), they are only useful for studying representations at the level of the general characteristics whereby the compared experimental conditions vary. Suppose, for instance, that during an fMRI experiment, subjects are presented with two sets of stimuli: (i) pictures of places and (ii) pictures of human faces. Experimenters may find out that a certain brain region or voxel responds differently to (i) as compared to (ii). However, from this univariate analysis, they can only infer that the target brain region is sensitive to the difference between places and human faces, that is, to the general dimension by which experimental stimuli can be disentangled.

The univariate approach does not allow us to infer how these neural activations associated with one set of stimuli (e.g., set (i)) differ from each other in terms of more specific stimulus features (e.g., whether the instances from set (i) are artificial or natural places); any decision as to which characteristics might be responsible for differences in activation is a matter of interpretation based on the researchers' hypothesis. Consequently, univariate approaches help us infer to what extent the general cognitive processing of one class of stimuli differs from another in neural terms, but cannot offer a deeper understanding of how stimulus features might be neurally represented. Indeed, for a richer analysis of brain activations associated with specific characteristics of stimuli, that is, their representations, we would need a method that can measure the multiple features or dimensions of experimental stimuli that might be processed by the brain. The limits of univariate analysis can be considered one of the main reasons why fMRI evidence has played such a minor role in the study of representations. In the next section we will consider the methods that have been introduced to address this issue.

2 Neural representations and multivariate methods

The neuroscientific study of representations has advanced in recent years, as new and more powerful methods have been proposed for analyzing neural activity. The most prominent of these is multivariate pattern analysis (MVPA).[13] MVPA differs from the univariate approach because it is based on the analysis of patterns of activations instead of single units and can potentially reveal a richer class of task-related effects.[14] Multivariate methods allow us to investigate how specific populations of neurons vary while processing stimuli and can help us understand how different voxels within the same region may process different information or features related to the experimental manipulation. Therefore, MVPA provides a multidimensional analysis of experimental conditions; it reveals the relationships between groups of voxels rather than mere voxel-by-voxel comparison. In other words, while the univariate approach assumes the single

voxel (or voxel cluster) as a whole encodes the manipulated psychological variable (e.g., processing places *vs* human faces), by analyzing different patterns of activation elicited across voxels, MVPA allows us to investigate how more specific stimulus features are neurally processed. Then, these patterns of activation detected by means of MVPA can easily be arranged within a "representational space" in which the responses of each of the considered voxels can be associated with specific features of the experimental stimuli.

Let us briefly describe the main steps involved in methods based on multivariate analysis. First, the acquired fMRI-based data are arranged in a N-dimensional representational space, showing changes in the neural activity for each of the N voxels considered over the time course of the study. Second, the data are split into two sets, the training and testing sets. The training set is labeled according to the experimental stimuli that generated the neural activity and used to train a machine learning algorithm (e.g., a support vector machine classifier). At this stage, the aim of the algorithm is to learn the best partition for the neural data, based on known experimental conditions. Third, once the algorithm has been trained on labeled data, it is tested using new, unseen data, that is, the testing set. The aim of this validation phase is to assess whether the partition previously established using the labeled data is also able to discriminate unseen data, that is, whether it accurately associates unseen neural activity with the experimental manipulation that generated it. Finally, if the algorithm succeeds in discriminating different populations of voxels, these patterns of voxels are considered to be responsive to the stimuli used during the experiment. Consider experimental stimuli sets (i) and (ii) above, respectively related to pictures of places and human faces. While univariate analysis would only allow us to disentangle faces and places in general, the representational space of neural activations based on MVPA would allow us to cluster stimuli within the same set (e.g., pictures of places) according to similarities (i.e., correlations) between their neural activations. At the cognitive level, such similarity is considered to provide a clue to neural processing of a certain feature (e.g., being a natural place). Thus, cluster of stimuli can be distinguished from other members of the same experimental set (e.g., those classified as pictures of artificial places). Therefore, MVPA offers a finer analysis of stimulus characteristics and allows us to analyze neural patterns in terms of similarity-based models, providing for more interesting connections between neural results and cognitive science theories.

According to advocates, multivariate analyses are more efficient than univariate analyses for investigating the problem of representation. This has lead to «representational studies»[15] in cognitive neuroscience, that is, studies that employ multivariate approaches to investigate how stimuli features are represented and neurally encoded in subjects' brains. However, it is worth noting that multivariate methods also suffer from general limitations; in many cases, they merely offer a more fine-grained answer to the problem of localization. Furthermore, the conclusions drawn by researchers using available fMRI data depend on the experimental designs used to collect such data (e.g., an ecological experimental paradigm as compared to a more classic design in which stimuli are presented repeatedly) as well as the methods[16] researchers selected to analyze them. Davis and Poldrack nevertheless highlight the advantages of representational studies that use specific multivariate methods, such as representational similarity analysis (RSA), when compared to the univariate approach.

RSA[17] represents one of the main methods for investigating the association between features of experimental stimuli and patterns of activation. The notion of similarity is central in this technique. Indeed, it allows researchers to assess to what extent stimuli sharing common characteristics can also be considered similar according to the neural activations they elicit.

Suppose that a researcher interested in affective neuroscience plans to investigate how humans categorize emotionally salient stimuli, such as pictures, in a fMRI setting. Assume that every picture presented during the experiment is characterized, among others, by a specific emotion-related feature like arousal (high or low). The experimental condition will then consist in manipulating the independent variable "arousal" by presenting different pictures[18] that should induce emotional responses with varying levels of arousal. In addition, the stimuli can be ordered according to high or low arousal; the researcher will expect that pictures which elicit a similar arousal level (e.g., high) will evoke more similar brain responses than pictures characterized by a different level of arousal (e.g., low).

RSA would allow us to analyze this type of isomorphism[19] between similar stimuli features and similar brain responses by (1) comparing stimuli pairwise within two distinct representational dissimilarity matrices (RDM), i.e., two matrices representing the degree of similarity of stimuli respectively assessed at the level of their neural activations and, for instance, behavioral responses, such as reaction-time, skin conductance, and self-report responses measured on specific scales;[20] and (2) assessing the overall, second-order similarity occurring between the two RDMs. Thus, RSA is one of the most powerful tools for running representational analyses because it assesses how stimuli feature and representational content with a dimensional structure (i.e., that vary according to one or more dimensions such as arousal) are neurally encoded. Moreover, it allows

us to detect similarities between multidimensional spaces built on different types of data, e.g., neural, behavioral, or derived from computational models, on the same set of experimental stimuli.[21]

3 Similar representations or common processes?

Neuroscientific representational methods allow us to directly address the problem of representation and other challenges in cognitive neuroscience, as pointed out by Davis and Poldrack:

> Given that one of the primary goals of cognitive neuroscience is to delineate how cognition is supported by the brain, it is highly desirable that neuroscience methods like neuroimaging be able to answer questions about both processes and representations.[22]

Assuming a clear distinction between the cognitive and neural levels of analysis of human cognition when trying to explain how neural representations are related to and influenced by cognitive representations and processes would represent a key step towards bridging the gap between theories in cognitive psychology and neuroscience.[23] So far, many works have already explored this line of research[24] and, within neuroimaging techniques, representational approaches seem to constitute the best way for studying how cognitive representations are neurally encoded. However, at this point, it is crucial to understand whether a potential gap between the cognitive and neural descriptions of the notions of process and representation may impact the discussion and the development of a mature cognitive neuroscience. A major concern is: What are representational methods applied to fMRI data actually picking up?

As Davis and Poldrack[25] note, this methodological issue is central when experimenters discuss whether fMRI data provides evidence for how the brain neurally represents stimuli characteristics or instead for how it implements a covarying cognitive process operating upon the same stimuli characteristics. In order to clarify this interpretational concern raised by Davis and Poldrack, we provide two toy-examples, describing two hypothetical fMRI-based investigations.

First, consider a fMRI-based MVPA study meant to investigate the cognitive process "face perception". Participants are presented with a battery of pictures representing several visual stimuli and asked to passively observe them during fMRI scanning. A subset of these stimuli are humans faces. In order to investigate face perception, the experimenters will analyze the brain activation patterns associated with viewing this subset of stimuli as compared to the rest. For the sake of simplicity, we will focus on two facial stimuli X and Y. Assume that, on average across participants, processing of X and Y show highly similar activation patterns, that is, activity of the same population of voxels. The experimenters can interpret these results for X and Y in two different ways, as summarized by the following interpretations:

> *(A1)*: The observed activation patterns are similar because both relate to processing the features of X and Y that are relevant for face perception (that is, a unique process operates on the common features of X and Y, such as their shape, the presence of a mouth, two eyes and the distances and positions of these eyes, that make face perception possible;
>
> *(B1)*: The observed activation patterns are similar because both relate to the representational similarity of X and Y (that is, the stimuli are representationally similar because both are facial stimuli and, therefore, share similar characteristics, such as the presence of a mouth, two eyes, their distances, and positions).

Even though apparently identical, the two interpretations of the neural activations are respectively associated with two distinct questions concerning what putatively happens at the cognitive level when stimuli are processed by subjects. Such questions can be formulated as follows:

> *(QA1)*: Which stimuli features are relevant for performing the process of face perception? (Connected to *A1*);
>
> *(QB1)*: Which features in the representations of the two stimuli are similar? (Connected to *B1*).

The point here is that the answer to *QA1* clearly overlaps with the answer to *QB1*. Indeed, we can expect that the same set of features in X and Y are relevant both for performing face perception and for assessing whether the two stimuli are representationally similar. In other words, performing face perception requires that X and Y be similar with respect to the fact that both are facial stimuli (e.g., have a certain shape, a mouth, two eyes, etc.).

In these scenarios the two questions are inseparable in practice and the two interpretations *A1* and *B1* about the neuroimaging evidence can be considered equivalent when analyzed at the functional level. Therefore, assuming that the neuroimaging evidence is shedding light on the first question (concerning the engagement of a certain cognitive process) rather than the second one (about the representations of stimuli), or *vice versa*, is simply impossible and, to some extent, even irrelevant for the interpretation of the results. In conclusion, it is not feasible (and probably meaningless) to ask whether similar neuroimaging evi-

dence for X and Y reflects the fact that participants are performing face perception or the fact that participants are processing the representational similarities of X and Y as facial stimuli.

Consider now a second toy-experiment in which neuroscientists are no longer interested in face perception, but rather in analyzing "emotion processing" and, in particular, the emotional responses elicited by stimuli. Suppose again that, for two stimuli X and Y representing human faces, similar populations of voxels are activated on average across participants. When asked to interpret this finding for X and Y, the experimenters may provide the following two interpretations:

> *(A2)*: The observed activation patterns are similar because X and Y trigger the same emotional response; (that is, a unique process operates on common affective features of X and Y that cause the same emotional response);
>
> *(B2)*: The observed activation patterns are similar because of the representational similarity of X and Y (that is, stimuli are representationally similar because both are facial stimuli and, therefore, share similar characteristics).

This second scenario may seem identical to the first one, but it is not. Again, the two interpretations can be associated with the following questions concerning the cognitive level:

> *(QA2)*: What stimuli features are relevant when stimuli are processed emotionally? (Connected to *A2*);
>
> *(QB2)*: Which features in the representations of the two stimuli are similar?" (Connected to *B2*).

However, in this scenario, given that we are interested in emotion processing rather than face perception, we would like to disentangle *QA2* and *QB2*. Indeed, the features of X and Y that make them similarly salient from an emotional viewpoint might not be identical to the features that make X and Y similar in other respects, e.g., as facial stimuli. In other words, when the functional level is analyzed, the interpretations *A2* and *B2* are not somehow equivalent as they are in the case of *A1* and *B1*, in which the overlap between face recognition as cognitive process and the similarity of X and Y as facial stimuli was not problematic. In this case, indeed, the feature(s) which permit X and Y to be grouped together in terms of emotional responses differ from the feature(s) according to which they might be judged to be representationally similar. Otherwise, it would mean that the general features characterizing a stimulus as a facial stimulus (e.g., the shape, the presence of two eyes, and the mouth) would be sufficient to cause the emotional response, leading to the same emotional response for an entire subset of facial stimuli, even when more specific aspects, such as the facial expression, changed dramatically. In this case, answering the two questions is not trivial and, in turn, the two interpretations *A2* and *B2* are not functionally comparable, even though the neuroimaging evidence on its own would make it impossible to disentangle them.

In both toy-examples, interpretations *A* rest on the investigation of the process that is commonly considered to be the input, while interpretations *B* rely on the representational similarity of the stimuli. For the *A*s, a similar pattern of activation would be evidence for concluding that the same cognitive process (e.g., emotional response) was engaged, while, for the *B*s, similar neural activations would be a clue to the processing of features that make the representations of those stimuli similar. However, the issue just described highlights how neuroscientists cannot completely disregard the possibility that neuroimaging results are not evidence of how the brain represents stimuli themselves, but may instead indicate what common cognitive process is recruited by the brain for operating on these representations.

The main difference between the two examples rests on the type of process that is considered: face perception or emotional response. Following a classic distinction proposed by Fodor,[26] Davis and Poldrack[27] argue that cognitive processes can be generally classified as domain-specific and vertical (e.g., vision) or domain-general and horizontal (e.g., categorization and decision-making). According to this distinction, the authors suggest that relating a similar activation pattern to representations rather than to a common process operating upon them is functionally equivalent only for domain-specific processes. Indeed, such processes appear to be confined to the elaboration of a particular feature that is common among stimuli of a specific type. On the contrary, domain-general processes appear able to operate over a wider range of types of experimental stimuli (e.g., emotional responses can be measured for pictures of facial stimuli as well as stimuli representing completely different contents such as movies) and systematically covary with some representational relationships between them.

Following these definitions, face perception might be classified as a domain-specific process, leading to the functional equivalence of *A1* and *B1*; while emotional response might represent an instance of horizontal processing, leading to non-functionally comparable interpretations, such as *A2* and *B2*. Therefore, the distinction between vertical and horizontal processing would allow neuroscientists to disentangle similarly problematic scenarios.

Specifying a priori the type of process putatively involved in an experimental investigation could

help address the interpretational limits of representational studies, but it is not enough to completely resolve the underlying conceptual issue. First, the distinction between vertical and horizontal processing is difficult in practice and may remain theoretically ungrounded if not embedded in a more general framework for human cognition. Second, it could be difficult for researchers to design experimental settings that completely control for the engagement of different types of processes or to establish – a priori – how these processes relate to each other and may impact the elaboration of stimuli.

Other strategies for accounting for these issues are available. A general strategy is using experimental designs in which all the potential features according to which stimuli vary and that can elicit the engagement of different cognitive processes are clearly distinguished. One way to implement this and, in turn, control for the cognitive processes recruited during the experimental manipulation, is to isolate the variable of interest by presenting classes of stimuli that have features that are orthogonal to the aspects being studied.[28] A further possibility is to analyze not only neuroimaging data, but also different types of evidence, such as behavioral data and results from computational models. This solution, available for instance in RSA-based studies,[29] can provide convergent information on the same set of stimuli, explaining the processes and representational features that participants of the experiment are actually handling.

To sum up, utilizing multivariate approaches for analyzing fMRI data is not enough to ensure an effective approach to studying cognitive representations. Specific strategies have to be taken to control experimental confounds, but these do not provide a definitive solution to a methodological issue that seems more conceptual than empirical. The relationships between cognitive processes and the neural evidence provided by neuroimaging techniques as well as the general use of the notions of representation and process in cognitive science and neuroscience need further theoretical clarification.

4 A cognitive ontology informed by Marr's "three levels of analysis"

The interpretational issues (and strategies to address them) discussed in the last section point towards a debate that has consequences for cognitive science as a whole, that is, the difficulty in providing an optimal "cognitive ontology". A cognitive ontology consists in a clear-cut taxonomy of mental states and cognitive processes recruited during task manipulation that can be efficiently mapped to available experimental evidence, such as brain activations. Many studies, at least in the neuroscientific field, have addressed these ontological aspects[30] and proposed tentative solutions, generally based on literature-mining and the creation of databases that include cognitive functions, tasks, behavioral variables, and neural evidence.[31] However, a major limit of these taxonomies is the coarse use of naïve and non-standardized labels for describing neuroscientific hypotheses in the literature. Moreover, these efforts appear to be mainly confined to the neuroscientific field, even though the problem of vague cognitive vocabulary affects several fields in cognitive science.

In order to better grapple with the interpretational issues that affect representational studies, we propose applying Marr's "three levels of analysis" as a general criterion for cognitive processes and representations. Indeed, Marr's analysis of the relationships between processes, representations, and physical implementations appears to provide a fine-grained vocabulary for describing our cognitive ontology and, in particular, promises to shed light on the vertical/horizonal processing distinction.

Marr[32] proposes to conceptualize the elaborations performed by a system as taking place at three levels: computation, algorithm, and implementation. The computational level describes what is computed and why, that is, the general logic of the computations performed by the system and their relevance for the task at hand. In particular, the computational level describes the computational model in abstract terms, that is, as a mapping between a given input and a computed output. Consider, for instance, a digital computer performing an arithmetical operation on numbers, e.g., a subtraction. The analysis of the computational level here will simply involve a description of the mapping between the input (e.g., a pair of numbers) and the output (e.g., a single number) and the set of rules and abstract properties that govern the operation at stake (e.g., the general properties of the operation of subtracting numbers).

The intermediate level is the algorithmic one, which describes how the computational theory is to be carried out. At this level, the notions of representation and process are pivotal, given that it is the algorithmic level that most closely depends on the system's architecture and how the system's architecture will determine how a certain input is represented and processed by the system itself. Representations consist in the entities storing information relevant for a certain task, that is, information about the available input from the external world. Processes refer to the algorithms that practically manipulate these representations. For instance, in the case of an arithmetical operation, representations may consist in numerical quantities we generally refer to by means of the symbols in the Arabic numeral system. Of course, we can use different notations for representing the same numerical entities, such as Chinese numerals: in this case we can talk about different representations. Therefore, the algorithms, i.e., the manipu-

lations that transform the input into the output of the process, largely depend on the features of these representations.

The algorithms for calculating subtraction based on the Arabic numeral system may differ from those based on another notation; the algorithms governing a digital computer will differ from those governing mathematical operations performed by means of an abacus. However, in all these cases, the algorithms will refer to the set of deterministic and finite operations that the system effectively performs on the internal representations to carry out the computation.

The last level of analysis discussed by Marr is the implementational one, involving the physical substrates with which the information processing is realized. For instance, a wooden abacus and a modern digital computer are two different physical implementations performing the same algorithmic operations. It is worth noticing that this level is independent from the previous one. Indeed, at least in principle, the same representational systems and algorithms can be implemented using different technologies. However, the characteristics of some algorithms might suit some substrates better than others.[33]

First of all, let us use Marr's model to reinterpret the toy-examples discussed in the previous section. Two preliminary clarifications are necessary. In Marr's model, the cognitive processes "face perception" and "emotional response" would refer to the general computations performed by the system, mapping certain stimuli provided as input (e.g., X and Y) to certain internal cognitive states given as output (e.g., an emotional reaction). Second, the neural activity observed during experiments would represent the fMRI-detectable trace of the implementational level, that is, what the fMRI scanner detects is assumed to be the neural realizer of the computation at issue. Consequently, Marr's model imposes a sort of linguistic shift: "face perception" actually refers to the general description of the computation at stake, without knowledge of how it is operationalized by specific processes operating on features of experimental stimuli, i.e., Marr's algorithms manipulating internal (cognitive) representations of the system. Moreover, the model suggests that the "neural representation", i.e., the fMRI-detectable neural activity, must now refer to physical realizers. Then, from now on, every time we talk about representations, we will specifically refer to the cognitive representations of external stimulus features on which cognitive processes, i.e., algorithms in Marr's terms, operate.

The interpretational issues discussed above arise at this level of analysis, that is, when we provide a description of how the computational level should be considered to be operationalized in terms of processes operating on representations. To clarify this aspect, let focus, for instance, on the computation "emotional response", from the second toy-example.

At the algorithmic level, we can suppose that what leads to the emotional response of the system is a series of processes, i.e., algorithms, operating on cognitive representations of the perceptual features of stimuli. In particular, such processes will extract some features from stimuli and then assess their emotional salience. This simple operationalization appears substantially consistent with interpretation *A2*. On the other hand, in this toy-example, interpretation *B2* is a potential confound. *B2* suggests that the same activation pattern (i.e., physical implementation) associated with the stimuli is actually related to their representational similarity. Therefore, we should assume that, in the case of *B2*, the computation "emotional response" is not operationalized by means of a series of processes operating on the representations of emotionally salient perceptual features. Instead, it should mark the representations of the stimuli themselves or, more correctly according to the assumed framework, a series of processes able to extract perceptual features that overall are common across stimuli and then to compare them in order to assess whether these stimuli are similar or not. Given that the emotionally salient features of X and Y are not necessarily equivalent to features that make them representationally similar (but covary with them), *A2* and *B2* are not functionally equivalent and, therefore, the interpretational issues arise.

The descriptions of computational theory at the level of the algorithms and the related representations may shed light on these problems. First, Marr's model allows us to disentangle the notions of representations and processes, by describing the interpretational issues only in terms of neural activations associated with different processes operating on representations. Second, Marr's model clarifies how we should conceptualize the difference between vertical and horizontal processing mentioned by Davis and Poldrack.[34] Indeed, it could be the case that one computation (e.g., face perception or processing of stimulus similarity) has to be considered vertical as compared to another (e.g., emotional response) when it is operationalized by algorithms able to take a lower number of types of representations as input, e.g., for face perception, only those referring to face-related perceptual features. In other words, a vertical computation appears operationalized by processes highly specialized for the elaboration of specific stimuli features, while a horizontal computation appears to be based on processes that can potentially operate on a variety of different representations. Therefore, two different cognitive functions can be considered comparable as vertical computations and also functionally equivalent, if the processes that im-

plement them are specialized for the elaboration of the same set of stimuli features. Instead, when a vertical computation such as the processing of stimuli similarity is compared with a more domain-general one, e.g., emotion processing (implemented by algorithms operating on a variety of different types of representations) their functional equivalence is no longer tenable. However, the features that make stimuli representationally similar systematically covary with those that make stimuli emotionally salient, leading to the interpretational issues discussed.

Of course, this type of analysis is always comparative, but the interpretational problem of representational studies always concerns two processes that need to be compared, e.g., the processing of stimuli similarity *vs* emotional response.

Thus, according to this classification, in the case of the two toy-examples discussed in 2, we can conclude that (i) the algorithms implementing face perception are as domain-specific as those related to the processing of stimuli similarity;[35] (ii) the algorithms implementing emotional response are more domain-general than those related to the processing of stimuli similarity; and, finally, (iii) algorithms implementing emotional response would also be more domain-general than those related to the processing of face perception.

Therefore, identifying Marr's algorithmic level allows us to disentangle different cognitive processes that remain apparently indistinguishable when described in more abstract terms, such as at the computational level. Indeed, the notions of algorithm and representation allow us to describe processes in fine-grained terms, that is, as elaborations operating on specific types of representations. The description makes clear what manipulations representations of stimuli features will undergo, so we can differentiate processes as algorithms that operate on different kinds of input. In this way, it clarifies the distinction between vertical and horizonal processing advanced in Davis and Poldrack[36] and improves the interpretation of results in representational studies.

A further advantage of Marr's model is that it allows us to specify the relationship between an experimental setting and the recruitment of cognitive processes. Indeed, without any information about the experimental setting, we cannot definitively rule out the engagement of one versus another cognitive process. However, differing from our toy-examples, experimental settings are usually described in considerable detail, for instance, detailing the instructions provided to subjects. The type of task performed and the instructions provided play a pivotal role in driving subjects' attention to certain features rather than others and, in turn, eliciting specific cognitive processes. For instance, in our second toy-example, the simple instruction to think about the emotional content of the presented stimuli may drastically change the cognitive performance of the subjects, by boosting their focus on emotionally salient characteristics. This means that an adequate cognitive ontology cannot avoid including a "task ontology"[37] in which experimental instructions, manipulations, and stimuli characteristics are clearly described and related to cognitive functions. However, to help us understand how a specific instruction may recruit a certain process that operates on specific stimuli features, we need a fine-grained description of cognitive states, not just a simple definition at the computational level. Marr's algorithmic level seems to offer a good framework. Moreover, it offers a principled motivation for explaining how experimental instructions supposedly relate to certain processes operating on specific representations of stimuli features.

5 Concluding remarks

The development of new techniques for analyzing brain activity has reopened the interesting debate on the impact of neuroscience in the field of cognitive science. Recently developed multivariate methods widely applied in cognitive neuroscience, such as RSA and machine learning analyses of fMRI data, are better tools than traditional univariate methods for shedding light on the notion of representation. This is a crucial challenge, since being able to explain how neural activations are related to and influenced by cognitive representations and processes would represent a key step in relating neuroscientific results to theories from cognitive psychology.

As we have argued, however, multivariate approaches are affected by issues related to the interpretation of neuroscientific evidence as a reliable clue to either cognitive representations of experimental stimuli or the covarying cognitive processes that operate on them. In order to clarify these issues on the methodological level, we detailed how they arise in the context of two toy-example fMRI-based representational studies. Then, we discussed the useful distinction between vertical and horizonal processing as well as other strategies for addressing this problem. We argued that these solutions only partially deal with the problem, since it appears to be more conceptual than empirical. Moreover, we noted that the vertical/horizonal processing distinction, as stated in Davis and Poldrack,[38] remains an insufficient instrument for mitigating the underlying methodological problems if not embedded within a more general theoretical framework.

In a nutshell, we claim that the main caveat in using multivariate methods to investigate cognitive representations is the lack of a unified theoretical framework that clarifies the relationships between representations, processes, and neural da-

ta, and also establishes guidelines using a fine-grained vocabulary, i.e., a cognitive ontology, that scientists can use to refer to cognitive states.

As a step towards better understanding concepts of representations and processes at the interface between cognitive science and neuroscience, we propose a novel application of Marr's well-known model to the methodological issues discussed above. More precisely, we suggest that Marr's model, by forcing researchers to describe how the computational level should be operationalized in terms of algorithms operating on representations, can shed light on the relationships between cognitive states and neural evidence.

Moreover, thanks to the notion of algorithm, the model disentangles cognitive processes that appear indistinguishable when described in only abstract terms, making distinctions such as the one between vertical and horizontal processing more clearly interpretable. Multivariate methods represent an important technique for analyzing human brain fMRI data and advancing cognitive science. However, their success crucially depends on how researchers deal with the methodological issues they raise regarding the relationships between cognitive representations, cognitive processes, and neural data. In this paper, we propose that approaching these problems from the perspective of cognitive ontology could improve theoretical dialogue comparing neuroscientific concepts and classical reflections in cognitive science. Indeed, a more accurate scientific taxonomy for human cognition is a first and necessary step towards a mature research program in cognitive neuroscience.

Acknowledgements and funding

The author would like to thank Fabrizio Calzavarini, Luca Cecchetti, Giacomo Handjaras, and Marco Viola for the discussions about the topics related to this work. A special thank to Gustavo Cevolani for the insightful suggestions in the development of the idea behind the current paper and the very helpful comments during the preparation of the draft. The author acknowledges funding from an ERASMUS+ Mobility Grant 2020/2021.

Notes

[1] Cf. M. Boden, *Mind as machine*; R. Núñez, M. Allen, R. Gao, C.M. Rigoli, J. Relaford-Doyle, A. Semenuks, *What happened to cognitive science?*.

[2] Cf. W. Bechtel, A. Abrahamsen, G. Graham, *The life of cognitive science*.

[3] Cf. R.A. Poldrack, *The physics of representation*.

[4] Cf. T. Davis, R.A. Poldrack, *Measuring neural representations with fMRI*.

[5] Cf. C.J. Price, K.J. Friston, *Functional ontologies for cognition*.

[6] Cf. D. Marr, *Vision*.

[7] Cf. A. Turing, *On computable numbers, with an application to the Entscheidungsproblem*; A. Turing, *Computing machinery and intelligence*.

[8] Cf. H. Putnam, *Psychophysical predicates*; D. Chalmers, *On implementing a computation*; R.L. Chrisley, *Why everything doesn't realize every computation*; J.A. Fodor, *The language of thought*. For a general overview of the different ways of construing the model, cf. G. Piccinini, C. Maley, *Computation in physical systems.*

[9] Cf. J.A. Fodor, *The mind-body problem*.

[10] Cf. D. Marr, *Vision.*

[11] A voxel is the minimal 3-dimensional unit for which the MRI scanner acquires the BOLD signal and produces images of the brain. The resolution of these images reflects the size of one side of the voxel, generally between 1 and 4 millimeters.

[12] Cf. T. Davis, R.A. Poldrack, *Measuring neural representations with fMRI*.

[13] Cf. J.D. Haynes, G. Rees, *Predicting the stream of consciousness from activity in human visual cortex*; K.A. Norman, S.M. Polyn, G.J. Detre, J.V., Haxby, *Beyond mind-reading*; N. Kriegeskorte, P.A. Bandettini, *Analyzing for information, not activation, to exploit high-resolution fMRI*; N. Kriegeskorte, M. Mur, P.A. Bandettini, *Representational similarity analysis-connecting the branches of systems neuroscience*; T. Naselaris, K.N. Kay, S. Nishimoto, J.L. Gallant, *Encoding and decoding in fMRI*; N. Kriegeskorte, R.A. Kievit, *Representational geometry*; J.B. Ritchie, D.M. Kaplan, C. Klein, *Decoding the Brain*.

[14] Cf. T. Davis, K.F., LaRocque, J.A. Mumford, K.A. Norman, A.D. Wagner, R.A. Poldrack, *What do differences between multi-voxel and univariate analysis mean?*.

[15] Cf. T. Davis, R.A. Poldrack, *Measuring neural representations with fMRI*; N. Kriegeskorte, P.K. Douglas, *Cognitive computational neuroscience*.

[16] For instance, as pointed out by a recent study that involved seventy neuroscientific research groups around the world, the specific pipeline used to analyze a single commonly available fMRI data set impacts the resulting conclusions and interpretations. R. Botvinik-Nezer, F. Holzmeister, C.F., Camerer, A. Dreber, et Alii, *Variability in the analysis of a single neuroimaging dataset by many teams*.

[17] In addition to MVPA, classification and machine learning methods can be used for multivariate analysis of fMRI data. The key aspect of machine learning representational approaches rests on training a model, such as a support vector machine to make classifications of neuroimaging data. Often these methods use brain activation patterns to determine to which experimental condition an unknown target stimulus belongs. A third type of representational analysis is encoding. Encoding models (cf. K.N. Kay, T. Naselaris, T., Prenger, J. L. Gallant, *Identifying natural images from human brain activity*; G. St-Yves, T. Naselaris, *The feature-weighted receptive field*; C. Du, J. Li, L. Huang, H. He, *Brain encoding and decoding in fMRI with bidirectional deep generative models*) can be thought of as working in the opposite direction to the classification approach: instead of mapping similarities between activation patterns onto stimuli features, encoding models map a hypothesized psychological feature space onto brain patterns (T. Davis, R.A. Poldrack, *Measuring neural representations with fMRI*, p. 118). RSA, machine learning, and encoding

can be considered to be different MVPA methods. However, for relevant relationships between them cf. *ibid*, p. 116; N. KRIEGESKORTE, P.A. BANDETTINI, *Analyzing for information, not activation, to exploit high-resolution fMRI*; N. KRIEGESKORTE, M. MUR, P.A. BANDETTINI, *Representational similarity analysis-connecting the branches of systems neuroscience*; N. KRIEGESKORTE, R.A. KIEVIT, *Representational geometry*; A.L. ROSKIES, *Representational similarity analysis in neuroimaging*.

[18] Assuming – as an anonymous Reviewer correctly pointed out – that there is a way to determine the level of arousal evoked by every picture that is univocal and independent from the observed neural activation.

[19] On the concept of isomorphism in cognitive science: R.N. SHEPARD, *Reviewed work*; R.N. SHEPARD, S. CHIPMAN, *Second-order isomorphism of internal representations*. On the concept of isomorphism in RSA, cf. N. KRIEGESKORTE, M. MUR, P.A. BANDETTINI, *Representational similarity analysis-connecting the branches of systems neuroscience*; N. KRIEGESKORTE, R.A. KIEVIT, *Representational geometry*.

[20] Suppose that, in addition to the fMRI study, experimenters collect self-report responses from a different group of subjects presented with the same set of pictures and asked to judge how emotionally salient these stimuli are on a scale from "no salience" to "maximum salience". These data will be useful for building one of the two RDMs, that is, the matrix comparing stimuli according to their affective characteristics. Therefore, such "behavioral" RDM will be compared to the other RDM, based on neural responses evoked by stimuli during the fMRI study.

[21] Cf. N. KRIEGESKORTE, M. MUR, P.A. BANDETTINI, *Representational similarity analysis-connecting the branches of systems neuroscience*; N. KRIEGESKORTE, R.A. KIEVIT, *Representational geometry*.

[22] Cf. T. DAVIS, R.A. POLDRACK, *Measuring neural representations with fMRI*, p. 108.

[23] In this work, we have discussed how multivariate methods can help shed light on the concept of representation and how a classic debate in cognitive science could be integrated with a more neuroscientific perspective. However, these methods do not offer any definitive answers to some deeper issues related to representations, for instance, the realism vs. anti-realism debate on mental entities and their nature (e.g., whether they are symbolic vs. distributed or multimodal vs. amodal). Multivariate methods do not represent a solution for these issues within the philosophy of cognitive sciences, but – we argue – as long as (a) a realist position on the concepts of representation and process and (b) a neuroscience-based perspective to studying the mind are assumed, multivariate methods can provide a more straightforward contribution to the study of representations and cognitive processes than other approaches to analyzing neuroimaging data.

[24] Cf. N. KRIEGESKORTE, P.A. BANDETTINI, *Analyzing for information, not activation, to exploit high-resolution fMRI*; N. KRIEGESKORTE, *Pattern-information analysis*; T. DAVIS, R.A. POLDRACK, *Quantifying the internal structure of categories using a neural typicality measure*; G. PICCININI, O. SHAGRIR, *Foundations of computational neuroscience*; C. BALKENIUS, P. GÄRDENFORS, *Spaces in the brain*; R.A. POLDRACK, *The physics of representation*.

[25] Cf. T. DAVIS, R.A. POLDRACK, *Measuring neural representations with fMRI*, p. 125.

[26] Cf. J.A. FODOR, *The modularity of mind*.

[27] Cf. T. DAVIS, R.A. POLDRACK, *Measuring neural representations with fMRI*, p. 125.

[28] Cf. *ibid.*; S. BRACCI, H.O. DE BEECK, *Dissociations and associations between shape and category representations in the two visual pathways*. However, as an anonymous Reviewer of this work correctly pointed out, it is worth noticing that the theoretical benefits of using certain experimental paradigms rather than others to disentangle different cognitive processes are often reduced or erased by similar difficulties that may arise at the empirical level.

[29] Cf. N. KRIEGESKORTE, M. MUR, P.A. BANDETTINI, *Representational similarity analysis-connecting the branches of systems neuroscience*; N. KRIEGESKORTE, R.A. KIEVIT, *Representational geometry*.

[30] Cf. C.J. PRICE, K.J. FRISTON, *Functional ontologies for cognition*; R.A. POLDRACK, *Can cognitive processes be inferred from neuroimaging data?*; R.A. POLDRACK, *Mapping mental function to brain structure*; C. KLEIN, *Cognitive ontology and region-versus network-oriented analyses*; J.C. FRANCKEN, M. SLORS, *From commonsense to science, and back*; M. VIOLA, *Carving mind at brain's joints*; T. EICH, D. PARKER, Y. GAZES, Q. RAZLIGHI, C. HABECK, Y. STERN, *Towards an ontology of cognitive processes and their neural substrates*.

[31] Cf. R.A. POLDRACK, A. KITTUR, D. KALAR, E. MILLER, C. SEPPA, Y. D. GIL, S. PARKER, F.W. SABB, R.M. BILDER, *The cognitive atlas*; J.L. LANCASTER, A.R. LAIRD, S.B. EICKHOFF, M.J. MARTINEZ, P.M. FOX, P.T. FOX, *Automated regional behavioral analysis for human brain images*; J.A. TURNER, A.R. LAIRD, *The cognitive paradigm ontology*.

[32] Cf. D. MARR, *Vision.*

[33] Cf. *ibid.*, p. 24.

[34] Cf. T. DAVIS, R.A. POLDRACK, *Measuring neural representations with fMRI*.

[35] Then, leading to the functional equivalence of interpretations A1 and B1 in the first toy-example discussed in section 4.

[36] Cf. *ivi*.

[37] Cf. C. FIGDOR, *Semantics and metaphysics in informatics*; R.A. POLDRACK, A. KITTUR, D. KALAR, E. MILLER, C. SEPPA, Y.D. GIL, S. PARKER, F.W. SABB, R.M. BILDER, *The cognitive atlas*; J.C. FRANCKEN, M. SLORS, *From commonsense to science, and back*.

[38] Cf. T. DAVIS, R.A. POLDRACK, *Measuring neural representations with fMRI*.

Literature

BALKENIUS, C., GÄRDENFORS, P. (2016). *Spaces in the brain: From neurons to meanings*. In: «Frontiers in Psychology», vol. VI, Art. Nr. 1820 – doi: 10.3389/fpsyg.2016.01820.

BECHTEL, W., ABRAHAMSEN, A., GRAHAM, G. (2017). *The life of cognitive science*. In: W. BECHTEL, G. GRAHAM (eds.), *A companion to cognitive science*, Blackwell, London/New York, pp. 1-104.

BODEN, M. (2006). *Mind as machine: A history of cognitive science*, Oxford University Press, Oxford.

BOTVINIK-NEZER, R., HOLZMEISTER, F., CAMERER, C.F., DREBER, A. ET ALII (2020). *Variability in the analysis of a single neuroimaging dataset by many teams*. In «Nature» vol. DLXXXII, n. 7810, pp. 84-88.

BRACCI, S., DE BEECK, H.O. (2016). *Dissociations and*

associations between shape and category representations in the two visual pathways. In: «Journal of Neuroscience», vol. XXXVI, n. 2, pp. 432-444.

CHALMERS, D. (1995). *On implementing a computation*. In: «Minds and Machines», vol. IV, n. 4, pp. 391-402.

CHRISLEY, R.L. (1995). *Why everything doesn't realize every computation*. In: «Minds and Machines», vol. IV, n. 4, pp. 403-430.

DAVIS, T., LAROCQUE, K.F., MUMFORD, J.A., NORMAN, K.A., WAGNER, A.D., POLDRACK, R.A. (2014). *What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis*. In: «Neuroimage», vol. XCVII, pp. 271-283.

DAVIS, T., POLDRACK R.A. (2013). *Measuring neural representations with fMRI: Practices and pitfalls*. In: «Annals of the New York Academy of Sciences», vol. MCCXCVI, n. 1, pp. 108-134.

DAVIS, T., POLDRACK, R.A. (2014). *Quantifying the internal structure of categories using a neural typicality measure*. In: «Cerebral Cortex», vol. XXIV, n. 7, pp. 1720-1737.

DU, C., LI, J., HUANG, L., HE, H. (2019). *Brain encoding and decoding in fMRI with bidirectional deep generative models*. In: «Engineering», vol. V, n. 5, pp. 948-953.

EICH, T., PARKER, D., GAZES, Y., RAZLIGHI, Q., HABECK, C., STERN, Y. (2020). *Towards an ontology of cognitive processes and their neural substrates: A structural equation modeling approach*. In: «PLoS ONE», vol. XV, n. 2 Art.Nr. e0228167 – doi: 10.1371/journal.pone.0228167.

FIGDOR, C. (2011). *Semantics and metaphysics in informatics: Toward an ontology of tasks*. In: «Topics in Cognitive Science», vol. III, n. 2, pp. 222-226.

FODOR, J.A. (1975). *The language of thought*, Harvard University Press, Harvard.

FODOR, J.A. (1981). *The mind-body problem*. In: «*Scientific American*», vol. CCXLIV, n. 1, pp. 114-125.

FODOR, J.A. (1983). *The modularity of mind: An essay on faculty psychology*, MIT Press, Cambridge (MA).

FRANCKEN, J.C., SLORS, M. (2014). *From commonsense to science, and back: The use of cognitive concepts in neuroscience*. In: «Consciousness and Cognition», vol. XXIX, pp. 248-258.

HAYNES, J.D., REES, G. (2005). *Predicting the stream of consciousness from activity in human visual cortex*. In: «Current Biology», vol. XV, n. 14, pp. 1301-1307.

KAY, K.N., NASELARIS, T., PRENGER, T., GALLANT, J.L. (2008). *Identifying natural images from human brain activity*. In: «Nature», vol. CDLII, n. 7185, pp. 352-355.

KLEIN, C. (2012). *Cognitive ontology and region-versus network-oriented analyses*. In: «Philosophy of Science», vol. LXXIX, n. 5, pp. 952-960.

KRIEGESKORTE, N. (2011). *Pattern-information analysis: From stimulus decoding to computational-model testing*. In: «Neuroimage», vol. LVI, n. 2, pp. 411-421.

KRIEGESKORTE, N., BANDETTINI, P.A. (2007). *Analyzing for information, not activation, to exploit high-resolution fMRI*. In: «Neuroimage», vol. XXXVIII, n. 4, pp. 649-662.

KRIEGESKORTE, N., DOUGLAS, P.K. (2018). *Cognitive computational neuroscience*. In: «Nature Neuroscience», vol. XXI, n. 9, pp. 1148-1160.

KRIEGESKORTE, N., KIEVIT, R.A. (2013). *Representational geometry: Integrating cognition, computation, and the brain*. In: «Trends in Cognitive Sciences», vol. XVII, n. 8, pp. 401-412.

KRIEGESKORTE, N., MUR, M., BANDETTINI, P.A. (2008). *Representational similarity analysis-connecting the branches of systems neuroscience*. In: «Frontiers in Systems Neuroscience», vol. II, n. 4 – doi: 10.3389/neuro.06.004.2008.

LANCASTER, J.L., LAIRD, A.R., EICKHOFF, S.B., MARTINEZ, M.J., FOX, P.M., FOX, P.T. (2012). *Automated regional behavioral analysis for human brain images*. In: «Frontiers in Neuroinformatics», vol. VI, Art. Nr. 23 - doi: 10.3389/fninf.2012.00023.

MARR, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information,* Henry Holt and Co., New York.

NASELARIS, T., KAY, K.N., NISHIMOTO, S., GALLANT, J.L. (2011). *Encoding and decoding in fMRI*. In: «Neuroimage», vol. LVI, pp. 400-410.

NORMAN, K.A., POLYN, S.M., DETRE, G.J., HAXBY, J.V. (2006). *Beyond mind-reading: Multi-voxel pattern analysis of fMRI data*. In: «Trends in Cognitive Sciences», vol. X, n. 9, pp. 424-430.

NÚÑEZ, R., ALLEN, M., GAO, R., RIGOLI, C.M., RELAFORD-DOYLE, J., SEMENUKS, A. (2019). *What happened to cognitive science?*. In: «Nature Human Behaviour», vol. III, n. 8, pp. 782-791.

PICCININI, G., MALEY, C. (2010). *Computation in physical systems*. In: E.N. ZALTA (ed.), *The Stanford encyclopedia of philosophy*–URL: https://plato.stanford.edu/entries/computation-physicalsystems/

PICCININI, G., SHAGRIR, O. (2014). *Foundations of computational neuroscience*. In: «Current Opinion in Neurobiology», vol. XXV, pp. 25-30.

POLDRACK, R.A. (2006). *Can cognitive processes be inferred from neuroimaging data?*. In: «Trends in Cognitive Sciences», vol. X, n. 2, pp. 59-63.

POLDRACK, R.A. (2010). *Mapping mental function to brain structure: How can cognitive neuroimaging succeed?*. In: «Perspectives on Psychological Science», vol. V, n. 6, pp. 753-761.

POLDRACK, R.A. (2021). *The physics of representation*. In: «Synthese», vol. CXCIX, n. 1, pp. 1307-1325.

POLDRACK, R.A., KITTUR, A., KALAR, D., MILLER, E., SEPPA, C., GIL, Y.D., PARKER, S., SABB, F.W., BILDER, R.M. (2011). *The cognitive atlas: Toward a knowledge foundation for cognitive neuroscience*. In: «Frontiers in Neuroinformatics», vol. V, Art. Nr. 17 – doi: 10.3389/fninf.2011.00017.

PRICE, C.J., FRISTON, K.J. (2005). *Functional ontologies for cognition: The systematic definition of structure and function*. In: «Cognitive Neuropsychology», vol. XXII, n. 3-4, pp. 262-275.

PUTNAM, H. (1967). *Psychophysical predicates*. In: W. CAPITAN, D. MERRILL (eds), *Art, mind, and religion*, University of Pittsburgh Press, Pittsburgh, pp. 37-48.

RITCHIE, J.B., KAPLAN, D.M., KLEIN, C. (2019). *Decoding the brain: Neural representation and the limits of multivariate pattern analysis in cognitive neuroscience*. In: «British Journal of Philosophy of Science», vol. LXX, n. 2, pp. 581-607.

ROSKIES, A.L. (2021). *Representational similarity analysis in neuroimaging: Proxy vehicles and provisional representations*. In: «Synthese», vol. CXCIX, n. 40, pp. 5917-5935.

SHEPARD, R.N. (1968). *Reviewed work: Cognitive psychology by Ulric Neisser*. In: «The American Journal

of Psychology», vol. LXXXI, n. 2, pp. 285-289.

SHEPARD, R.N., CHIPMAN, S. (1970). *Second-order isomorphism of internal representations: Shapes of states*. In: «Cognitive psychology», vol. I, n. 1, pp. 1-17.

ST-YVES, G., NASELARIS, T. (2018). *The feature-weighted receptive field: An interpretable encoding model for complex feature spaces*. In: «Neuroimage», vol. CLXXX, pp. 188-202.

TURING, A. (1936). *On computable numbers, with an application to the Entscheidungsproblem*. In: «Proceedings of the London Mathematical Society», vol. XLII, pp. 230-265.

TURING, A. (1950). *Computing machinery and intelligence*. In: «Mind», vol. LIX, n. 236, pp. 433-460.

TURNER, J.A., LAIRD, A.R. (2012). *The cognitive paradigm ontology: Design and application*. In: «Neuroinformatics», vol. X, n. 1, pp. 57-66.

VIOLA, M. (2017). *Carving mind at brain's joints. The debate on cognitive ontology*. In: «Phenomenology and Mind», vol. XII, pp. 162-172.