Moscow State University of Psychology & Education

Vladimir Kozunov

Interaction of value-dependent and sensorimotor mechanisms of sensory feature integration in holistic visual perception of an object

Dissertation Summary
for the purpose of obtaining academic degree
Doctor of Philosophy in Psychology

Academic supervisor:
Tatiana Stroganova
Prof., Doctor of Science

Moscow - 2021

Table of contents

The dissertation was prepared at the Center for Neurocognitive Research (MEG Center), Moscow State University of Psychology and Education.

Publications:

1. Kozunov, V. V., & Ossadtchi, A. (2015). GALA: group analysis leads to accuracy, a novel approach for solving the inverse problem in exploratory analysis of group MEG recordings. *Frontiers in neuroscience*, *9*, 107. DOI: 10.3389/fnins.2015.00107

2. Kozunov, V., Nikolaeva, A., & Stroganova, T. A. (2018). Categorization for faces and tools - two classes of objects shaped by different experience - differs in processing timing, brain areas involved, and repetition effects. *Frontiers in human neuroscience*, *11*, 650. DOI: 10.3389/fnhum.2017.00650

3. Kozunov, V. V., West, T. O., Nikolaeva, A. Y., Stroganova, T. A., & Friston, K. J. (2020). Object recognition is enabled by an experience-dependent appraisal of visual features in the brain's value system. *NeuroImage*, *221*, 117143. DOI: 10.1016/j.neuroimage.2020.117143

Selected Conferences:

- XVII All-Russian conference "Neuroinformatics-2015", Moscow, Russia, 2015

- IHNA&NPh RAS - Conference «Integration of information in the visual analyzer: from a feature detector to a holistic visual image», dedicated to the 85th anniversary of the birth of Academician I.A. Shevelev, Moscow, Russia, 2017

- VIIIth International Conference on Cognitive Science, Svetlogorsk, Russia, 2018

- All-Russian scientific conference in memory of J. Bruner "Psychology of cognition: low-level and high-level processes", Yaroslavl, Russia, 2020

- CAICS-2020 - I National Congress on Cognitive Research, Artificial Intelligence and Neuroinformatics, Moscow, Russia, 2020

- MEG NORD 2021 International Conference in MEG Research, Moscow, Russia, 2021

Introduction

One of the main tasks of psychology is to analyze the principles of formation and development of mental phenomena. The solution to this problem is inextricably linked with the study of physiological processes underlying mental manifestations. In this context, the key is the question of the emergence of "units of analysis" [Vygotsky, 1999], which have irreducible properties of mental phenomena.

During visual perception, a picture of the surrounding world is formed, consisting of a set of isolated objects. Holistic images of objects, characterized by their meaning at the conceptual level, are units of visual perception, which introduce the specificity of mental phenomena representation. The problem of the emergence of these units is partly determined by the fact that the flow of light on the retina is not divided into separate objects, and external influences are so complex and variable that there is an infinite number of variants for their perceptual organization. The study of integration processes, during which an emergent property - conceptual (semantic) meaning - is formed, is an important task, the solution of which will answer a number of fundamental questions of psychology. Meanwhile, despite the fact that the isolation of separate features of visual stimuli has been studied quite well, the criteria and mechanisms of their integration for holistic perception of an object still remain unclear.

In structuralism (Wundt and Titchener), which followed the ideas of associationism in philosophy, it was believed that perception is constructed from atoms of elementary, unrelated local sensations. The units of perception are synthesized as a result of the work of two mechanisms - the summation of isolated sensations and the association of the sum with memory image. The problems of structuralism were mainly caused by the impossibility of explaining the perceptual constancy when changing local elements and, accordingly, with the inadequacy of the summation law as a universal principle of integration.

A possible way to solve this problem was indicated by von Ehrenfels [Ehrenfels, 1937]. Noticing that perceptual experiences, such as the perception of melody or the shape of a visual object, are more than the mere sum of their independent components, he postulated a new sort of properties - Gestalt quality (Gestaltqualität). Ehrenfels's Gestalt qualities were superadded to our experiences of sensory elements and existed alongside them. In the Berlin school of Gestalt psychology, the view of holism was more radical. Rejecting the premise that the sum of sensory elements constitutes the primary foundation of perceptual experience, Wertheimer [Wertheimer, 2007] objected to any summative account in which something is added to the sum of sensory elements, be it von Ehrenfels gestalt quality, Helmholtz's unconscious inferences [Helmholtz, 1866] or conscious mental operations imposed on sensory elements to produce unity. He argued that we directly and immediately perceive Gestalt: integrated, structured wholes the properties of which are not derived from its individual parts but that the parts themselves can only be defined in relation to the whole.

Having pointed out that Gestalt is capable of "transposition", Gestalt psychology postulated the presence of functional structures, independent of sensory content, that determine the forms of perceptual organization. However, the a priori introduction of Gestalt principles leaves unanswered the question of the nature of perceptual structures. An attempt to reduce the origin of Gestalts to the existence of universal laws of organization ("physical forms") was unsuccessful, as such laws imply, at least, their invariance in the course of mental development. J. Piaget [Piaget, 1947] was one of the first to show the existence of age-related evolution of mechanisms, culminating in the formation of perceptual invariants.

Piaget proposed to consider perception as "the product of a progressive construction which arises, not from "syntheses", but from adaptive differentiations and combined assimilations" [Piaget, 1947]. He borrowed from the "Thought Psychology" the concept of an "anticipatory schema" - a complex of preliminary relations, in which "the solution of a problem cannot be reduced to the stimulus-response schema, but

consists of filling in the gaps in complexes". As a result, perception was presented as "filling in" the anticipatory schema with the "relations" complementing the "complex". And the problem of holistic perception appeared in the form of a question about the relationship between the processes of emergence of anticipatory schema and the processes of implementing the corresponding relations. Piaget suggested that the anticipatory schema originates from the equilibrium of assimilating structures, which arises not in connection with a particular process of differentiation of the visual flow, but is associated with "personal space and time, a scale of values, etc." [Piaget, 1947] and persists for a long time.

In modern cognitive psychology, the idea of anticipation has been developed in detail in terms of computational algorithms. Among them, two models should be mentioned: predictive coding [Rao, Ballard, 1999] and biased competition [Vecera, 2000]. Both theories suggest a hierarchy of processing levels and an interaction between top-down expectations and bottom-up sensory signals. However, predictive coding assumes that downstream influences suppress neural activity by the amount predicted by higher cortical regions so that only the (residual) error between prediction and upstream signals propagates from one cortical region to the next. On the contrary, the model of biased competition assumes that signals ascending from stimuli increase a neural activation in the neural circuitry of expectations consistent with them, thereby biasing the competition occurring between expectations in favor of one of them (or a limited number of them).

A key feature of predictive coding theory is the need for a formalized rule(s) by which higher levels generate predictions at lower levels - the so-called forward model [Friston, 2005]. On the one hand, this provides the theory with mathematical clarity and makes it especially attractive to cognitive science, traditionally resting on the information processing approach. On the other hand, the price of the "information-theoretical elegance" of predictive coding is its limitation to the principles of structuralism and the inheritance of its conceptual problems. Indeed, "sufficient statistics", in terms of which predictions are formulated (generated by the forward

model), are computed on the basis of regularities in the stimulation flow according to the same summation law (with an inessential, in this case, addition of normalization).

An alternative to the explicit representation of the predicted signals (and their direct comparison with ascending from stimulus ones) can be an anticipatory mechanism in which sensory signals are assimilated by consistent steady states of neural activity. Consistency, in this case, corresponds to synergetic action within the functional system [Anokhin, 1973] and is based not on the similarity between the "template" and the stimulus, but on the complementarity (adding what is missing), which assists in performing the function of the system. The result of assimilation of sensory signals, which corresponds to "filling in the gaps that exist within anticipatory schema", is a bias of the competition between different functional states in favor of a limited (small) number of them. Such a biased competition model costitutes an implicit anticipatory mechanism, which does not require representations of the expected states of the system. Anticipation, in the sense of realizing distal (or higher level) states, emerges as the result of the coupling of the dynamics at different scales [Pezzulo, 2008]. At the global level, the system is coupled with the environment into the close cycle, and the steady-state of their combined dynamics determines the functional system itself.

However, a system without representations, involved in a closed cycle with the environment, is only able to respond to external changes. The ability for voluntary behavior supposes the existence of decoupled from the environment, substitute representations that can be manipulated even in the absence of stimuli. Representations so defined must have two contradictory characteristics: autonomy (from the cycle with the environment) and intentionality (directedness toward something in the environment). Indeed, autonomy from behavior presupposes abstract representations that lack meaning (and, as a consequence, justification of the origin) - this is the symbol grounding problem [Harnad, 1990].

G. Edelman [Edelman GM, 1992] proposed as a solution to this problem (although Edelman himself avoided the term "representation" and used, depending on the

context, the semantically close terms "concept" and "scene") to divide the organism-environment cycle into two relatively independent parts, which correspond to two types of organization of the nervous system: the value system and the thalamocortical system. The value system - mainly the limbic system massively associated with the endocrine and autonomic nervous systems - differentiates environmental influences on the basis of their accordance to meet the needs of a given organism (i.e., on the basis of subjective value). Relatively independent of this, the thalamocortical system categorizes external events on the basis of sensorimotor associations ("perceptual categorization" in Edelman's terms). A distinctive feature of the value system (which is a consequence of its close connection with systems that maintain homeostasis) is its partial autonomy from the environment and the ability to "categorize its own activities" [Edelman G.M., 1992]. As a result, the brain can create and maintain structures (categories of conceptual memory) derived from subjective appraisals, but devoid of their direct connection with the ongoing moment. In the proposed model of perception, the conceptual memory interacts with brain areas carrying out ongoing perceptual categorizations of world signals, resulting in a "scene" (representation) as a correlate of the phenomenological image of consciousness.

At the conceptual level, this model corresponds to Piaget's idea - value-dependent structures assimilate relations between spatial features. However, Edelman suggested the "Neural Darwinism" theory to "show in detail how both perceptual and conceptual categorization can arise as a result of [neural group] selection" [Edelman, 1993]. This opened a way for explaining the mechanism of the emergence of abstract anticipatory schemes. Complementarity originated within one functional system, can be ("horizontally") transferred to another if both of them are subsystems of a more global system at the underlying - neural - level of organization. As a result, a mechanism can arise in which "coherence of this scene [representation] is coordinated by the conceptual value-category memory even if the individual perceptual categorization events that contribute to it are causally independent" [Edelman G.M., 1992]. Thus, the autonomy of representations is provided by the initial divergence of value-dependent

and sensorimotor categorizations, while intentionality is a consequence of their later convergence during the formation of a conscious image. The fact that during convergence the integration of spatial features is only partially determined by causality in the external environment (captured by perceptual categorization) is a prerequisite for the emergence of a new (mental) level of behavior determination.

We propose to introduce additional differentiation into the Edelman model, which in particular will account for the differences in the perception of natural and man-made categories of objects. It is known, that the thalamocortical pathways (at the level of the visual cortex) can be divided into two relatively independent subsystems: the ventral and dorsal visual streams. It was suggested [Goodale, Milner, 1992] that performing various functions, these pathways form a visual subsystem "for representation" (ventral) and a subsystem "for action" (dorsal). By reconciling these views with Edelman's theory (and with neuroanatomical data), it can be assumed that these subsystems differ in their decoupling from the sensorimotor cycle (by varying degrees of mediation using value-dependent structures). This determines their different autonomy degree, which is a necessary attribute of representations.

The dorsal system, lacking representations, is often associated with Gibson's theory [Gibson, 1979], in which motor affordances determine perception. The concept of affordances, however, is not well defined. While it was for a long time used to refer to (an almost unlimited) set of actions or manipulations that are simply given by spatial properties of the environment (e.g., if you have a chair and a ball in front of you, then you can kick both of them in the same way, as well as sit on them). Another more recent strand of research deals with affordance as the (rather limited) set of action options that emerge from experience and learning what an object is made for (an affordance for a chair is to sit but not kick, and vice versa for a ball). There are three points to be made about this. First, both interpretations of affordances describe them as a set of motor actions (in space). Secondly, in the first interpretation, the set of actions cannot, and in the second, it can quite accurately determine the category of the object.

Finally, the more artificial (man-made) object is considered, the more meaning the second interpretation has.

Based on these remarks, two types of sensorimotor associations can be distinguished (which approximately corresponds to the hypothesis of two separate subsystems: "reaching to move" and "reaching to use" - [Daprati, Sirigu, 2006]) depending on what is considered as "sensory" part. Low-level associations link motor schemes with simple spatial features (statistics), that are free from value-dependent influences originated from conceptual memory. Such features are necessary for the representation of any categories as they provide spatial "content" to value-dependent anticipatory schemes. High-level associations bind motor schemes with previously represented "parts". Such associations determine the ways of integration for the perception of objects that were artificially created as sets of parts functionally combined according to the mental level rules (originated in the depths of individual conceptual categorization, but necessarily distributed among individuals within a society). In contrast to the complementarity-based appraisals in the value system, the combined parts are represented (encoded), therefore they can be directly compared with the (also represented) predictions generated by the forward model. Accordingly, the operation of this mechanism can be described by a predictive coding model.

The presented theory of perception assumes that the main criterion for integrating features into a holistic image of an object is the characteristics of the experience that led to the isolation of a particular category of objects and determine their meaning. As a consequence, one expects that in the perception of different categories, two integration mechanisms - value-dependent and sensorimotor - will be involved to different degrees. Previously, specific spatial patterns of neural responses were found for many categories including faces, animals, houses, places, tools, artificial objects, etc. [e.g., Bracci, Beeck Op de, 2016; Ishai et al., 2000]. However, experiments in this area were mainly carried out using functional magnetic resonance imaging (fMRI), which did not allow determining the temporal parameters of processes that caused pattern differences. Accordingly, it was problematic to make a conclusion about what

causes them - preconscious processes at the stage of feature integration or later processes.

In my dissertation research, I investigated the emergence of spatio-temporal patterns of neural activity specific to perception of holistic images of two categories. For this, magnetoencephalography (MEG) - a neuroimaging method with a high temporal resolution - was used. In the first part of the work, a new toolkit for the analysis of MEG data based on multivariate pattern analysis (MVPA) is presented. I developed this toolkit to obtain localization maps of the brain sources of experimental effects with millisecond time resolution, which is required for the study of separate stages of such short processes as visual perception.

In the experimental study (parts 2 and 3), I applied the paradigm, during which the subjects were presented with bitonal images (Mooney figures) [Mooney, Ferguson, 1951] of objects of two categories. These categories were chosen to be on the opposite ends of a spectrum relating to the prevalence of either value-dependent or sensorimotor experience of interactions with objects of these categories. Some of these images were degraded in such a way that when presented to a "naive" subject, they did not evoke the perception of a meaningful image. However, after the training procedure, which includes the demonstration of the original non-degraded image, the subjects recognized a meaningful object. This paradigm made it possible to compare the conditions when unrecognized and recognized as a meaningful object visual stimuli were completely identical in terms of their sensory characteristics and differed only in the subject's experience.

On this methodological basis, in the second part of the work, I tested two hypotheses concerning the differences in the integration processes during the perception of objects of either natural or artificial categories. The first hypothesis was that for the dorsal mechanism to start binding of separate parts based on high-level sensorimotor associations, the ventral mechanism had to finish performing its function and had formed preliminary representations of parts. Accordingly, we expected that the

activity in the dorsal pathway, which is specific for the processing of functionally defined objects, would manifest itself later than the face-specific activity associated with the enriched processing in the ventral pathway.

The second hypothesis concerns the dynamics of the activity of value-dependent and sensorimotor mechanisms during the facilitation of object recognition caused by repetitive presentations of the same stimulus. Visual stimulation plays different roles for these mechanisms: either it is assimilated by a dynamic anticipatory scheme or it informs about a prediction error (these two cases are described by models of biased competition and predictive coding, respectively). So one should expect that the facilitation of face recognition is accompanied by an increase in the amplitude of face-specific evoked activity in the ventral pathway. On the contrary, learning high-level sensorimotor associations leads to a decrease in the computational burden caused by the task of eliminating the error between predictions and actual representations, and therefore the amplitude of tool-specific activity will decrease.

In the third part, I tested the hypothesis that rapid, experience-dependent appraisal of visual signals in the brain's value system can pre-select anticipatory schemes. The problematics of this issue is determined by the fact that, in most cases, the neural codes for a combination of spatial features extracted from stimulation are characterized by degeneracy, i.e., structurally different codes represent the same feature combination or, vice versa, one code represents different combinations. Such degenerative coding is not specific enough to select a conceptual category, and a narrowed set of anticipatory schemes must be preselected [Pizlo, 2001]. The variant of selection presented by our hypothesis is an alternative to that proposed previously [Bar et al., 2006]. The latter selection principle is based on faster processing of low spatial frequencies, which provides a global structure organizing the integration of slowly processed high spatial frequency features. The value-dependent mechanism is more general as it solves the problem of selection of conceptual, i.e. abstracted from spatial characteristics (including global), anticipatory schemes. As a confirmation of this hypothesis, we expected during each act of perception to register activity in the

cortical regions of the value system and predicted that this activation can be modulated by experience, i.e., by recognition of a degraded image induced by a specific procedure. Crucially for the value-based selection hypothesis, the formation of category-specific patterns in the high-level visual cortex for newly recognized meaningful images should follow the activation of the brain value system.

Key aspects to be defended:

1. In order to study the microgenesis of a holistic percept, an instrumentarium is needed for binding data on brain activity with the stages of deployment of psychological processes. The method of MEG data analysis - RB-MVPA - provides an ability for simultaneous determination of high-resolution temporal parameters and localization of activity in the brain, which helps to relate experimentally recorded data with characteristics of psychophysiological processes underlying mental phenomena.

2. Two feature binding mechanisms, differing in either value-dependent or sensorimotor criteria of integration, determine holistic perception of a meaningful object. The value-dependent mechanism in the ventral subsystem is functioning during the perception of any category of objects but its activity is more pronounced when visual features allow representing a natural category enriched by value-dependent characteristics. In the perception of functionally defined categories, later in time, the sensorimotor mechanism of the dorsal visual pathway is involved in a specific way.

3. The integration mechanisms of the ventral and dorsal visual subsystems have opposite dynamics of evoked activity changes accompanying multiple repetitions of stimuli. These changes are respectively caused either by more efficient assimilation by the corresponding anticipatory scheme of visual signals (ventral) or by a decrease in the prediction error (dorsal).

4. During experience-induced recognition, reorganization of visual cues appraisals in the value system of the brain precedes an appearance of the categorical structure of neural activity in the high-level areas of the visual cortex. This suggests that object recognition is enabled by experience-dependent appraisals of the visual cues in the

value system (salience map formation) that induce a pre-selection of anticipatory schemes required for binding distinct aspects of visual input into a coherent percept.

<h1 style="text-align:center">Overview of the study</h1>

Part 1. Elaboration of methods of MEG data analysis for studying the processes of visual perception in the human brain

From the moment a visual stimulus appears to the emergence of a conscious image of an object, 250-300 ms passes. To observe separate stages of such a rapid process, it is necessary to use methods that allow detecting changes in brain activity with millisecond resolution. Encephalography has such a temporal resolution, however, in many cases, does not allow to localize temporal effects in the brain with the required accuracy. The problem is caused by the fact that measurements in encephalography are made outside the brain, and localization of the activity in the cerebral cortex must be found in the process of solving the inverse problem. To solve an ill-posed inverse problem, it is necessary to use model assumptions that limit the range of solutions and, ultimately, allow one to choose a unique solution.

In the study [Kozunov, Ossadtchi, 2015], we proposed a new method for solving the inverse problem - GALA (Group Anatomy Leads to Accuracy). In this method, the relaxed requirement of similarity of activations in different subjects in a group study was used as the main model assumption. In GALA, we, first of all, establish equivalence between sources of activity in different subjects in the anatomical sense. To do this, we, using Freesurfer segmentation results, associate each vertex of the cortex presented in the form of triangulation mesh of one subject with one (or a linear combination, depending on the selected options) vertex of another subject (and so for all subjects). To express mathematically the model assumption, used as priors for solving the inverse problem, we design a special structure of vertex by vertex covariance matrices. This structure expresses the requirement that similar (functionally equivalent) activations in different subjects should be localized in close, but not necessarily coinciding (anatomically equivalent) vertices of the triangulation mesh.

The correctness of the solution to the inverse problem drastically depends on the physiological plausibility of the model assumptions. We found that GALA provides a significant increase in the accuracy of localization of neural activity when a high degree of coincidence of functional and anatomical equivalence of individual vertices can be expected (for example, when studying responses in retinotopic regions of the cortex). However, when using GALA in higher-level regions of the cortex, the results are similar to those obtained using the simplest algorithm for minimizing the quadratic norm [Hämäläinen, Ilmoniemi, 1994]. The obvious reason for this is the implausibility of the requirement for (even relaxed) vertex-to-vertex similarity of activities in different subjects when studying high-level processes.

In the study [Kozunov, Nikolaeva, Stroganova, 2018], we proposed an alternative approach to obtain sufficiently accurate spatial maps of experimental effects. Method RB-MVPA is based on multivariate pattern analysis (MVPA) borrowed from machine learning. In this approach, we first solve the inverse problem in a common way. After that, we group activities of vertices corresponding to the large-scale (containing at least 50 vertices) regions aforehand selected on the basis of the anatomical atlas. The basic idea behind this approach is to compare (based on the accuracy of the MVPA classification) experimental conditions in a neural activity feature space of an individual subject and conceptualize differences in multiple vertices as a single value. Only following the transition to this scalar indicator of between-condition differences in brain activity we evaluate the similarity between subjects by performing statistical analysis. As a result, on the one hand, we abandon the requirement for vertex-to-vertex consistency of activities in different subjects, extracting experimental effects only at the large scale level, where the coincidence of functional and anatomical similarities between subjects is a more physiologically plausible assumption. On the other hand, we retain the ability to detect changes in the internal structure of the activity of the regions with an accuracy of each vertex (or the number of principal components representing the activity in the region).

As shown in [Kozunov, Nikolaeva, Stroganova, 2018], the RB-MVPA method allows one to obtain spatial maps of experimental effects that are in good agreement with those obtained using fMRI, however, with a millisecond time resolution. This makes this method unique and allows it to be successfully applied to study stages of visual perception.

Part 2. The study of distinguishing characteristics of integration processes during perception of faces and tools

In the experimental part of the study [Kozunov, Nikolaeva, Stroganova, 2018], we investigated successive stages of visual stimulation processing when perceiving objects from a natural category - faces, and an artificial category - the tools. We tested the hypothesis that, within the stage of integration of spatial features, the activity specific to the artificial category (in the dorsal subsystem) will manifest itself later than for the natural one (in the ventral one). In addition, we studied the differences in the dynamics of activities in the ventral and dorsal subsystems accompanying perceptual learning during multiple repetitions of stimuli.

Twenty-two subjects (10 men, 12 women) - mean age 25.4 years (SD = 4.62) participated in an experiment in which we were presenting bitonal images (Mooney figures) and asking the subjects to answer which object they saw in the picture. The images depicted objects from categories of faces, tools, animals, plants, as well as might not contain meaningful objects. In this part, we analyzed only three classes of images: containing faces, tools, and meaningless images. The stimuli were displaying to the participants using Presentation software via a computer with a 60 Hz frame rate and were back-projected on a translucent white projection screen located 1.7 m in front of the participants to provide an 8 × 8 degrees visual angle. Each image (of the above classes) was presented 160 times in a pseudo-random order (mixed with the rest of the classes to reduce expectation of a stimulus from a particular category) with a duration of 800 ms and an interval between stimuli pseudo-randomly varying from 1000 to 1500 ms. The stimuli presentation was divided into two sessions (two blocks

each), and the responses in each of them were analyzed as two separate sets (to generalize the results).

During the demonstration of images, neuromagnetic responses were recorded with the helmet-shaped 306-channel sensor array (Neuromag). Continuous data was divided into epochs from -500 ms prior to stimulus onset, until 1000 ms post-stimulus. Following the selection of epochs, the timeseries data were low-pass filtered at 24 Hz and baseline corrected using a -300 to -1 ms interval before stimulus onset. Then we transferred the data into the source space of cortical activity, for which we applied an anatomically constrained inverse problem solver, forcing the sources to lie on a tessellated mesh of the cortical mantle. The meshes for every participant were obtained on the basis of individual high-resolution structural T1-weighted MRIs acquired on a 1.5 T Toshiba ExcelArt Vantage scanner.

Further analysis of MEG data was carried out using the RB-MVPA method described in Part 1 and its supplement - representational similarity analysis (RSA, [Kriegeskorte, Mur, Bandettini, 2008]). RSA was used to interpret pairwise classification results obtained using RB-MVPA. We have composed three types of model representational dissimilarity matrices (RDM) to extract experimental effects. The first type corresponded to the model of low-level spatial features processing of different stimuli. In this model, high differences in each combination of pairwise comparisons of different stimuli were contrasted with low differentiation of the same stimuli in the two sessions. The second type was used to highlight the activity that characterizes the isolation of a meaningful category - low rates of differences within classes, one of which corresponded to a meaningful category and the second consisted of meaningless images, were contrasted with high rates of differences between classes (the model was applied separately for faces and tools - pairwise comparisons with the group corresponding to the alternative meaningful category were not considered). RDMs of the common structure of the third type were used for two purposes: 1) to highlight the "supra-categorical" activity, i.e. processing of meaningful stimuli without taking into account their categorical specifics; 2) to highlight specific activity for a particular

meaningful category. In these RDMs, as in the model of the second type, low differences within classes were contrasted with high differences between classes. However, pairwise comparisons for all groups (two meaningful categories and meaningless images) were considered simultaneously so that two of them were combined into one class. To highlight the "supra-categorical" activity, two meaningful categories were combined into one class. To highlight specific activity for a particular meaningful category, an alternative meaningful category was combined with a meaningless images group. The similarities between empirical (based on pairwise comparisons of MEG data computed with RB-MVPA) and model RDMs were assessed by calculating Spearman's correlation coefficients.

Using the sequential application of RB-MVPA and RSA, we experimentally determined the temporal boundaries and spatial localization of three stages of visual processing in the brain, corresponding to 1) low-level processing, 2) category-specific integration of spatial features, and 3) "supra-categorical" processing. Spatio-temporal patterns of activity specific for isolation of the category of faces were found in the areas of the ventral visual pathway, and only at the stage of categorization in the time window of 140-180 ms. In contrast, tool-specific activity was found both at the categorization and later stages of processing. The tool-specific activity of most interest, during the feature binding stage, was found in the interparietal sulcus (IPS) of the left hemisphere (high-level region of the dorsal visual pathway) in a time window of 210-220 ms, i.e., as expected, appeared later than face-specific activity.

In addition, we found that face-specific activity in the ventral regions increased during perceptual learning with multiple repetitions of the same stimuli. In contrast, tool-specific activity in the dorsal subsystem decreased during presentation of repetitive stimuli. Our results indicate the existence of two mechanisms of integration of spatial features, localized in different regions of the brain and with different dynamics of neural activity underlying stimuli learning.

Part 3. The study of the influence of value-dependent experience on spatial features integration processes

In the study [Kozunov et al., 2020], we tested the hypothesis that under the influence of the previous experience, visual signals entering the value system are re-estimated, and their appraisal enables the emergence of a categorical structure of activity in the high-level visual cortex leading to perception of a meaningful object.

Thirty-four subjects (15 males, 19 females) - mean age 24.6 years (SD = 4.31) - participated in the experiment in which we were presenting bitonal images (Mooney figures) and asking the subjects to answer which object they saw in the picture. The stimuli and the method of their presentation were similar to those described in Part 2. However, in this study, we mainly analyzed the processing of images (not considered in Part 2), perception of which was dependent on the fast training procedure designed to induce stimulus recognition. These images: "naively unrecognizable faces" and "naively unrecognizable instruments", were specially designed to satisfy the following requirement: They should be identified as meaningless by more than 90% of subjects, when seen for the first time but should be correctly identified as an object (face or tool) by more than 90% of subjects when seen after performing a training procedure. Testing and selection of such stimuli were carried out in aforehand procedure on 60 neurotypical volunteers, none of whom participated in the main experiment.

Registration and preprocessing of MEG data also followed the steps described in Part 2. Additionally, we used audio recordings of the subjects' responses to select only those trials in which the subjects "correctly" identified an image, i.e. before training as a meaningless one, and as relevant objects after training. In the course of the experiment, the subjects could "make mistakes": either see a meaningful object (not necessarily a face or instrument) before training or do not see a face or a tool after training. We rejected all such trials, as well as entire classes of images if there were less than 50 correct trials before or after training.

As in Part 2, data analysis was performed using RB-MVPA and RSA. In this study, we used a model RDM to highlight the effects caused by the transition to recognition of a meaningful object in a picture, which before the training was perceived as a set of meaningless blots. This RDM entailed a higher dissimilarity between activity patterns for stimuli across recognized-unrecognized group boundary than within groups. The first group consisted only of the class of "naively unrecognizable faces/tools" after training (i.e. recognized), while the second combined "naively unrecognizable faces/tools" before training with the class of meaningless images (in both sessions), i.e. every meaningless images. In addition, we applied two methods based on the same decoding strategy used in RB-MVPA. In order to investigate the possibility that some of the stages may be shifted in time during processing of different classes of objects, we used the temporal cross-category generalization analysis. We also distinguished the processing phases in which decoding of conditions was caused either only by an increase in the total power of a region activity or by reorganization of the activity within the region without a change in power of the whole region. To this aim, we analyzed the internal structure of the MVPA classification weights and differentiate a proportional increase in the amplitudes of all components of the MVPA feature space without changing their relative weights from a change of the weights without changing the total region's power.

Recognition induced changes were first seen in the 100-120 ms window after stimulus onset in the extrastriatal regions of the right hemisphere, for both faces and tools. However, these changes were not categorically specific and were characterized by an increase in the total power of the activity pattern that existed before learning (i.e., determined by low-level spatial features). This profile of changes indicates that they were, most likely, caused not by a recognition-inducing procedure but by gradual perceptual learning that accompanies multiple repetitions of stimuli.

A quite different type of changes in brain activity was found in the 210-230 ms window in the areas of the brain's value system. In the category-dependent regions of brain's value system, namely, insula, entorhinal cortex, and anterior cingulate of the

right hemisphere for faces and right orbitofrontal cortex for tools, patterns of neural activity demonstrated reorganization of activity (within the regions) without an overall increase in power in these regions. This effect emerged with a concomitant differentiation between exemplars within the group of recognized images (contrasted to the absence of such differentiation for the same stimuli before training). That is, the after-training patterns of activity in the regions of the value system, most likely, was not associated with a certain conceptual category, but demonstrated an altered and only partially similar structure of appraisal of different (for separate instances within the category) spatial cues.

Further, using temporal cross-category generalization, we found that during the perception of induced recognition face stimuli, patterns of activity in the area of the right fusiform gyrus are formed, which are distinguished by the same classifiers that were used to decode the perception of "simply recognizable faces" (recognized without the need for a training procedure). This face-specific response emerged at 240–290 ms, with a much greater latency than during perception of simply recognizable faces, and, crucially, followed the reorganization of activity effect in the value system regions. That is, in the case of experience-induced perception, one can clearly distinguish between early appraisal in the value system and later selection of category-specific anticipatory schemes, that characterized by the emergence of face-specific activity in the right fusiform gyrus. During the induced perception of tools, we did not register the emergence of tools-specific patterns in the dorsal subsystem. This is consistent with the fact that our training procedure led to a change in the appraisal of visual cues in the value system and did not change the sensorimotor schemes responsible for the generation of tool-specific activity. Apparently, the conscious recognition of the instruments was induced by the altered (more "correct") representation of the image's parts.

Key Results and Conclusions

Key results:

1.   The RB-MVPA methodology, which was developed here, allows linking the experimentally recorded parameters of MEG signals to the distinct stages of processing hierarchy underlying the holistic perception of two classes of visual objects whose recognition was shaped either by value-dependent (faces) or functionally defined sensorymotor experience (tools). Using RB-MVPA we have extracted three stages of cortical activity corresponding to low-level processing, category-specific feature binding, and supra-categorical processing of the Mooney figures of faces and tools and gained evidence of their modulations induced by recognition of these figures as the meaningful objects.

2.   Feature binding mechanisms that integrate the separate features of visual input into a coherent object operates at relatively early stage of visual processing – 130 – 240 ms after the image onset, and show strikingly selective spatial-temporal cortical patterns in response to simply recognizable faces and tools. The perceptual categorization is followed by later stage of mentalizing-related processing common for both meaningful categories, that starts at 250 ms and includes widely distributed assemblies within parietal, temporal and prefrontal regions.

3.   At the feature binding stage of simply recognizable figures, the face-specific spatiotemporal patterns are associated with the bilateral activation of ventral occipito-temporal areas at 140-170 ms, while for tools the selective activity is delayed up to 210-220 ms and occurs in the dorsal visual pathway, which enables an access to motor schemes associated with the earlier decoded representations of the "parts" of visual object.

4.   The integration mechanisms of the ventral and dorsal visual subsystems have opposite dynamics of activation accompanying multiple repetitions of the respective images – repetition enhancement of the selective neural response for faces and its repetition suppression for tools. The augmentation of neural responses to repeatedly

presented face images may correspond to more effective assimilation of visual signals by the appropriate anticipatory schema, which causes its selection facilitation. Whilst repetition suppression/habituation for tools is, most probably, related to reduced functional load on the sensorimotor mechanism to explain away a prediction error.

5. The perception of "naively unrecognizable" figures after their induced recognition evinced category-specific reorganization of activity within 210-230 ms time window in the regions of the brain's value system (insula, entorhinal cortex and cingulate of the right hemisphere for faces and right orbitofrontal cortex for tools) that preceded the selective activity (240-290 ms) in the higher-tier visual areas (e.g. right fusiform gyrus for faces) involved in feature binding. We hypothesized that the top-down signal from value system to the visual areas enables category-specific processing to occur.

6. New patterns of activity in the areas of the value system during perception of "induced recognition" images allow better decoding between different examples of stimuli within-category found for both face- and tool- related cortical regions at the feature binding stage. That is, these patterns in the areas of the value system, most likely, were not associated with anticipatory schemes of semantic categories, but are changed (after training) saliency maps of distinct visual features.

7. The results suggest that perceptual synthesis in the high-level areas of the visual cortex is enabled by an experience-dependent appraisal of visual cues in the value system of the brain that prompts the selection of the anticipatory scheme required for binding distinct aspects of visual input into the coherent percept.

Scientific novelty and significance of the results.

We have developed two new methods for analyzing MEG data. In GALA, we for the first time proposed using the variability of the measured data for individual subjects (which usually makes analysis difficult) as an additional source of information that reduces uncertainty in solving the ill-posed inverse problem and improves localization accuracy. The result of this study was later used for new algorithms for solving the

inverse MEG problem [Janati et al., 2020]. In the RB-MVPA method, we implemented a new approach to the analysis of the spatiotemporal structure of brain activity based on the application of the MVPA algorithm to the activities of sources within the boundaries of anatomically selected regions (the full set of which covers the entire cortex). To this aim, we solved the problem (arising due to the fact that the transition to the source space in the analysis of MEG data is performed by solving an ill-posed inverse problem) of the balance between preserving location specificity and sufficient dimensionality of the data feature space to discriminate between experimental conditions.

Until now, localization of activity specific to visual perception of various categories was obtained mainly using fMRI, which did not make it possible to isolate the processes at the conscious and preconscious stages of perception. With the help of MEG and RB-MVPA, we investigated the spatiotemporal structure of brain activity during holistic perception of objects and for the first time presented direct evidence of the involvement of the dorsal pathway mechanism in the preconscious processes of feature integration during perception of functionally defined objects.

Our results allow significant progress in understanding the mutual role of two different criteria for the integration of spatial features into a holistic image of an object. The existence of two subsystems in the brain - ventral and dorsal - as well as their differential involvement in the perception of natural and man-made objects has been known for a long time [Mishkin, Ungerleider, Macko, 1983; Goodale, Milner, 1992]. However, the question of how the two mechanisms of integration - value-dependent and sensorimotor - interact with each other and determine an inner representation of a unique meaningful object, has not yet been clarified. We have shown for the first time that the process of preconscious integration of features is divided into two substages, characterized by the sequential activation of value-dependent and sensorimotor subsystems, differing in the dynamics of changes accompanying multiple repetitions of stimuli.

These results substantiate a new model of integrative processes in visual perception. Selection and representation of anticipatory schemes based on a value-dependent criterion in the ventral subsystem is a prerequisite for the perception of any categories of objects. These processes are more pronounced for natural categories of objects, the richness of value-dependent characteristics of which determines a greater potential for informative selection of schemes. The dorsal mechanism starts to act after a representation of a limited set of anticipatory schemes is completed. It tries to bind (if possible, as in the case of tools) intermediate representations - "parts" - on the basis of functional (high-level sensorimotor) experience. The successive, but fast (in particular, not distinguishable by fMRI) launching of two subsystems with different dynamics of processes, allows one to explain a number of controversial data on brain mechanisms of recognition facilitation [for a review, see Segaert et al., 2013].

Our results confirm the importance of the thesis about the "unity of affective and cognitive processes" for solving the fundamental issue of the formation of units with irreducible properties of mental phenomena inherent in perception. We found that value-dependent (affective) appraisals of stimulation are performed in the brain not independently of cognitive processes of perception, but allow selection of anticipatory schemes, which is required for the formation of an image of a meaningful object. Results partially leading to such a conclusion were presented earlier. In particular, Adolphs [Adolphs, 2002] has introduced the two pathways model for the appraisal of value-based aspects of a stimulus: a subcortical path is specialized for very fast, automatic extraction of characteristics of stimuli that convey intrinsic, biologically relevant values, whilst along the cortical path, complex visual features are feed into a network of frontal regions to recognize characteristics invoking personal salience and conveying values based on the extent of personal associations learned over past experiences. In the fMRI study [Ludmer, Dudai, Rubin, 2011] it was shown that the activity of the structures of the brain value system during the training procedure was a predictor of the subsequent recognition of degraded images. In our study, we showed for the first time that the reorganized structure of appraisals in the value system

actually manifested itself in each act of perception and preceded the integration of features in the high-level visual cortex.

Finally, our results provide a previously missing argument for clarifying the question: "What makes the perception of faces special?" A number of specific effects, which characterize the holistic nature of face perception, are described in numerous psychological studies. To explain this specificity, many researchers claim the existence of a specialized module located in the right fusiform gyrus of the cerebral cortex for the perception of faces. Their opponents, whose point of view do we share, argue that face-specificity is determined by a set of attributes dependent on value experience. The evidence favoring the latter view mainly relies upon the capacity of so-called expert-level experience with the use of any objects to create the unique holistic perception of their visual images both on the behavioral [e.g., Boggan, Bartlett, Krawczyk, 2012] and at the physiological levels [Gauthier et al., 2000; McGugin et al., 2014]. Our findings on the delayed activation of the fusiform gyrus during the perception of degraded faces complemented these previous studies and strengthened the value-based account. We have shown, for the first time, that the face-specific activation of the fusiform gyrus appeared only after the top-down feedback from the value system to visual cortex, which enables high levels of processing to form category-specific anticipatory schemes.

The relevance of the topic of dissertation research is associated not only with the fundamental scientific problems of studying the principles of organization and implementation of mental activity but also with the practical significance of the results. Perceptual problems are closely related to our understanding of the nature of subjectivity and voluntary action. Their solution will open the way to the elaboration of fundamentally new technologies for preventing erroneous human actions caused by the so-called "human factor". Another important area of application of new knowledge about brain integration mechanisms is the development of technologies for the rehabilitation of people with impaired perception functions after brain damage as a result of trauma, cerebrovascular accidents, and other pathological processes,

including mental illness. In the future, new knowledge about the reorganization of the functional structures of the brain under the influence of experience can and should be used to create more effective educational programs, including for people whose learning ability is altered as a result of developmental disorders or neurodegenerative diseases.

References

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews*. https://doi.org/10.1177/1534582302001001003

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmidt, A. M., Dale, A. M., … Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 449–454. https://doi.org/10.1073/pnas.0507062103

Boggan, A. L., Bartlett, J. C., & Krawczyk, D. C. (2012). Chess masters show a hallmark of face processing with chess. *Journal of Experimental Psychology: General*, *141*(1), 37–42. https://doi.org/10.1037/a0024236

Bracci, S., & Op de Beeck, H. (2016). Dissociations and associations between shape and category representations in the two visual pathways. *Journal of Neuroscience*, *36*(2), 432–444. https://doi.org/10.1523/JNEUROSCI.2314-15.2016

Daprati, E., & Sirigu, A. (2006). How we interact with objects: learning from brain lesions. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2006.04.005

Edelman, G. M. (1993). Neural Darwinism: Selection and reentrant signaling in higher brain function. *Neuron*. https://doi.org/10.1016/0896-6273(93)90304-A

Edelman G.M. (1992). *Bright air, brilliant fire: On the matter of the mind.* Basic books.

Ehrenfels, C. V. (1937). On Gestalt-qualities. *Psychological Review*, *44*(6), 521–524. https://doi.org/10.1037/h0056968

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Gauthier, Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*(2), 191–197. https://doi.org/10.1038/72140

Gibson, J. J. (1979). The Ecological Approach to Visual Perception. *Houghton Mifflin-Boston*.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*. https://doi.org/10.1016/0166-2236(92)90344-8

Hämäläinen, M. S., & Ilmoniemi, R. J. (1994). Minimum-Norm Estimation. *Medical & Biological Engineering & Computing*, *32*(1), 35–42.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1–3), 335–346. https://doi.org/10.1016/0167-2789(90)90087-6

Helmholtz, H. von. (1866). *Treatise on Physiological Optics*. *Book*. Retrieved from http://poseidon.sunyopt.edu/BackusLab/Helmholtz/

Ishai, A., Ungerleider, L. G., Martin, A., & Haxby, J. V. (2000). The representation of objects in the human occipital and temporal cortex. *Journal of Cognitive Neuroscience*, *12 Suppl 2*, 35–51. https://doi.org/10.1162/089892900564055

Janati, H., Bazeille, T., Thirion, B., Cuturi, M., & Gramfort, A. (2020). Multi-subject MEG/EEG source imaging with sparse multi-task regression. *NeuroImage*, *220*. https://doi.org/10.1016/j.neuroimage.2020.116847

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

Ludmer, R., Dudai, Y., & Rubin, N. (2011). Uncovering Camouflage: Amygdala Activation Predicts Long-Term Memory. *Neuron*, *69*(5), 1002–1014. https://doi.org/10.1016/j.neuron.2011.02.013

McGugin, R. W., Newton, A. T., Gore, J. C., & Gauthier, I. (2014). Robust expertise effects in right FFA. *Neuropsychologia*, *63*, 135–144. https://doi.org/10.1016/j.neuropsychologia.2014.08.029

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*. https://doi.org/10.1016/0166-2236(83)90190-X

Mooney, C. M., & Ferguson, G. A. (1951). A new closure test. *Canadian Journal of Psychology*, *5*(3), 129–133. https://doi.org/10.1037/h0083540

Pezzulo, G. (2008). Coordinating with the future: The anticipatory nature of representation. *Minds and Machines*, *18*(2), 179–225. https://doi.org/10.1007/s11023-008-9095-5

Piaget, J. (1947). *The Psychology of Intelligence. The Psychology of Intelligence*. https://doi.org/10.4324/9780203278895

Pizlo, Z. (2001). Perception viewed as an inverse problem. *Vision Research*, *41*(24), 3145–3161. https://doi.org/10.1016/S0042-6989(01)00173-0

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Segaert, K., Weber, K., de Lange, F. P., Petersson, K. M., & Hagoort, P. (2013). The suppression of repetition enhancement: A review of fMRI studies. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2012.11.006

Vecera, S. P. (2000). Toward a biased competition account of object-based segregation and attention. *Brain and Mind*, *1*(3), 353–384. https://doi.org/10.1023/A:1011565623996

Wertheimer, M. (2007). Gestalt theory. In *A source book of Gestalt psychology.* (pp. 1–11). https://doi.org/10.1037/11496-001

Anokhin, P. K. (1973). Principial'nye voprosy obschej teorii funkcional'nyh sistem. In *Principy sistemnoj organizacii funkcij.* (pp. 5–61).

Vygotsky, L. S. (1999). *Myshlenie i rech'*, 1999. 323 p.