Filozofia Nauki
Rok XXIII, 2015, Nr 2(90)

Błażej Skrzypulec

Relational Construction of Visual Objects*

It is widely recognized in cognitive sciences that stimuli which affect the visual system are quite different from what we consciously see (e.g. Palmer 1999). The perceptual system is stimulated by light waves, but our visual field is usually filled with objects. It seems that the visual system must be able to construct representations of objects by relying on the input information.

Within this context, one can ask what concept of object could adequately describe the way in which such representations of objects, or "visual objects", are formed. Psychological and neuroscientific models often consider processes that allow construction of visual objects (e.g. Treisman 1999). Unfortunately, they usually fail to make all their conceptual assumptions and consequences explicit, nor do they sufficiently explain which philosophical concept of object is used in their framework.

On the other hand, various accounts of the structure of objects have been proposed by philosophical theories of individual objects, formulated in the framework of analytic metaphysics. Most philosophical debates revolve around notions offered by the substratum (e.g. Martin 1980), bundle (e.g. van Cleve 1985), and substantial (e.g. Lowe 2006) theories. However, scientific theories concerning vision are rarely addressed in these metaphysical discussions, and the proposed accounts of the object structure were not usually applied to visual objects.

In this article, I consider examples of philosophical investigations concerning the structure of visual objects that can be found in works by Austen Clark (2004) and Athanassios Raftopoulos (2009) as well as assumptions present within scientific models of vision, especially those concerning figure/ground discrimination (e.g.

* The research for this paper was supported by the National Science Centre, Poland (grant no. 2012/05/N/HS1/03408). The author would like to thank Adam Chuderski and Claudio Masolo for their comments regarding an earlier version of the paper.

Vecera 2000). In reference to these accounts, I analyze how non-object visual representations are transformed into visual objects. I argue that feature-placing mechanisms (proposed by Clark) as well as segmentation and grouping processes are insufficient to fully explain the way in which visual objects are constructed, even if they are combined with Pylyshyn's notion of visual indices (Pylyshyn 2001). However, in some works by Raftopoulos (e.g. 2009: 7-15) and in scientific models of figure/ground discrimination, the visual object construction is characterized in terms of competition between visual regions. Such statements may be interpreted as supporting the relational view on visual object formation, according to which proto-object structures become visual objects in virtue of standing in certain external relations to other, nearby structures. My main goal is to present a more precise formulation of the relational view regarding the visual object construction which is implicitly accepted in scientific models of vision and to argue that it offers a better approach to describing the structure of visual objects than standard philosophical — bundle, substratum, and substantial — theories of objects.

I start with conceptual considerations about the understanding of key notions such as "visual object" and "constructing characteristics". Firstly, I propose a criterion for distinguishing visual objects from other types of visual representations. Secondly, I explicate the notion of "constructing characteristics". Then I investigate at which stage of the perceptual process visual objects are formed. I describe contemporary scientific models associated with that stage and bring out their conceptual assumptions concerning visual object construction. Finally, I test compatibility of the most popular contemporary philosophical notions of object structure with the assumptions of the scientific models under consideration. I conclude that relational factors are the most relevant in the context of visual object construction.

1. CONCEPTUAL CONSIDERATIONS

1.1. Representation and visual objects

In this paper, a visual object is characterized as a representation (Palmer uses the term "object-based representation" — 1999: 91). The perceptual system relies on the information received on the retina and constructs models — visual objects — which represent the source of the incoming stimuli. Sometimes the term "visual object" is understood differently as denoting things that are not models/representations but external sources of stimuli which are represented as objects by the visual system (e.g. spatially coherent material beings). Below I refer to those external, represented sources of stimuli as "physical objects" or "elements of the environment".

If we seek a more precise account of visual objects, we should distinguish three elements connected with the notion of "representation":

(1) Physical object — visual objects represent external, physical objects which are sources of stimuli received by the perceptual system. Those physical objects are not identical with visual objects, but it may be the case that under ordinary conditions the structure of visual objects quite accurately represents the structure of physical objects.

(2) Vehicle of representation — incoming information for the visual system is processed by physical structures that are parts of the neural system. These devices may be called "vehicles of representation", yet they are not visual objects, and their structure (layers of neurons connected in a certain way) is not the structure of visual objects.

(3) Visual representation — in order to represent physical objects, the visual system constructs models of external reality by using retinal information processed by neural devices. Some of these models represent fragments of the perceiver's environment as objects. I refer to such models as "visual objects" (a term "object-based representation" is also in use, see Palmer 1999: 91) and investigate what type of structure they possess.

The difference between the three notions can be demonstrated by a simple example. It may be the case that physical objects are represented by the visual system as localized bundles of features (size, colour, etc.). In such a situation, the structure of visual objects consists of at least two distinct kinds of elements (i.e. location and feature[1]) combined by some relation. This structure does not necessarily match the structure of the external, physical object and is clearly different from the structure of the neural mechanisms that contributed to the creation of the visual object.

1.2. Which representations are visual objects?

Visual objects are not the only "inhabitants" of the visual space. There are also various pre-object representations — such as edges, blobs, regions constituting ground behind figures, etc. (Marr 1982, Palmer, Rock 1994b). Thus a conceptual problem arises: what is the criterion for deciding whether a certain type of visual representation is a visual object? It is easy to recognize developed visual objects in our usual visual experiences: when we have conscious access to the final product of the visual system, we perceive some parts of the visual field as occupied by three-dimensional, coherent, complex individuals. It is much harder to decide what, if anything, serves as a visual object at the earlier stages of the perceptual process.

A common way of avoiding this problem is to make an abstraction from the features of developed visual objects. With such an approach, more primitive visual ob-

[1] It should be noted that here "location" and "feature" do not refer to physical places and features but are labels for simpler representations which constitute the structure of a more complex one — the visual object.

jects share some characteristics with the fully developed ones: for example, they represent fragments of the environment as spatially coherent (Spelke 1990) or as being able to persist through change (Raftopoulos, Müller 2006a: 254), though they lack some of their features (they are not sufficient to recognize something as a member of a general category). Unfortunately, the choice of the relevant features will always be arbitrary to some extent, and it is quite probable that all visual representations, including the most primitive ones, share some features with fully developed visual objects. Furthermore, the above approach hardly investigates what the visual system treats as visual objects — rather, it exploits our intuitions about what it means to be an object.

In this paper, I choose a different criterion. I do not start by deciding whether a type of visual representation is a visual object based on features shared by its tokens but rather by determining the position of that type within the perceptual process. This criterion relies on the notion of "developed visual representation" that may be intuitively accepted as a paradigm example of a visual object:

(*DVR*) A developed visual representation is a type of visual representation whose tokens represent fragments of the perceiver's environment as entities that are exemplars of general categories, have a definite shape (often, but not always, in the form of a complex 3D structure), and are accessible within the standard visual awareness.

Different, more primitive visual representation types are related to DVR by way of some process. The relation which is especially relevant here may be called "rigid connection":

(*RC*) A type of visual representation is rigidly connected to DVR iff every token of that type is transformed into a DVR token when it is processed non-regressively.

The processing is "non-regressive" when it does not degrade the representation's structure by turning it into a more basic one. Regressive processes occur when there is a shortage of time or operational capacities, for example, when attention is rapidly drawn to another part of the visual field or when the presentation of stimuli is very brief.

Some types of visual representations may not be rigidly connected to DVR, which means that, even under very good processing conditions, at least some of their tokens are not transformed into DVR tokens.

Based on (*DVR*) and (*RC*), we can define the notion of visual object as follows:

(*VO*) A visual object is a type of visual representation which is identical with DVR or which is rigidly connected to DVR.

The main advantage of the above definition is that it seems to successfully cut out a special fragment of perceptual processing, in which all visual representations are processed in a way that transforms them into paradigm examples of visual objects, i.e. DVRs. It allows us to determine what counts as a visual object by considering the way visual system is functioning, without having to rely on our intuitions about what it means to be an object.

According to (*VO*), many types of visual representations may be visual objects. These types form a hierarchy which has DVR as its highest element. At the bottom of the hierarchy are the most primitive types of representation which are still rigidly connected to DVR, that is to say, all tokens of those types are transformed into DVR tokens under good processing conditions.

In what follows, I investigate, by referring to scientific models, which stage of visual processing gives rise to representations that belong to the types which are rigidly connected to DVR. To avoid confusion, let us adopt a terminological convention: in order to refer to types of visual representation capital letters will be used (e.g. Visual Object), while expressions referring to their tokens will be typed in lower case (e.g. visual object).

1.3. Constructing characteristics

So far I have presented my understanding of the term "visual object" and proposed the criterion for distinguishing visual objects from different types of visual representation. Based on that, I can explain what I mean by "visual object construction".

The criterion (*VO*) decides whether representations are tokens of a Visual Object by considering their place within the perceptual process. In the subsequent sections of the paper, I try to figure out on what structural grounds representations gain the status of visual objects, i.e. by virtue of what structural characteristics they become tokens of Visual Objects in the sense given by (*VO*).

I assume that visual objects have some structure which is described by a set of characteristics, for example, "is composed of two kinds of elements" etc. I also assume that tokens of a single type of visual representation share a common set of structural characteristics (CSSC for short). The question remains how to decide which of the many structural characteristics connected with types that are Visual Objects are those by virtue of which visual objects are constructed.

In order to answer this question, we may introduce the notions of "minimal visual object" and "maximal non-object representation":

(*MVO*) A type of visual representation is a Minimal Visual Object iff

(1) it is a Visual Object, and

(2) there is no subset of the CSSC shared by the tokens of that type which is a CSSC shared by the tokens of a different type that is also a Visual Object.

In other words, the visual representation type is a MVO when there is no other type with a "smaller" CSSC — in the sense of being a subset — which is still a Visual Object. It is worth noting that there may be several types of representation that satisfy the definition (*MVO*). The notion of "maximal non-object representation" can be defined as follows:

(*MNOR*) A type of visual representation is a Maximal Non-object Representation iff

(1) it is not a Visual Object, and
(2) there is an extension of the CSSC shared by the tokens of that type which is a CSSC shared by the tokens of a type that is a Minimal Visual Object, and
(3) there is no other type of visual representation such that it is not a Visual Object and the CSSC shared by its tokens needs smaller (in the sense of being a subset) expansion to become a CSSC shared by tokens of a Minimal Visual Object.

Simply speaking, a type of visual representation is a MNOR when its CSSC is the closest one to the CSSC of a type that is a Minimal Visual Object. Similarly to the case of Minimal Visual Objects, the definition (*MNOR*) does not exclude a situation in which several types of representations are Maximal Non-object Representations.

By using the notions of MVO and MNOR, it is possible to characterize the "constructing" structural characteristics for visual objects:

(*CON*) A set of structural characteristics is constructing iff it constitutes a difference between the CSSC shared by the tokens of a Minimal Visual Object and the CSSC shared by the tokens of a Maximal Non-object Representation.

Visual object construction may be characterized as a process in which maximal non-object representations are transformed into minimal visual objects by gaining constructing characteristics. My goal is to find out which structural characteristics may be identified as constructing and which types of ontological elements they describe.

1.4. Structure of objects in philosophy

The framework of analytic metaphysics has been dominated by three general ideas about the structure of objects. I will label the first concept "unique element structure" (UE). According to this notion, entities are objects only when they possess a unique element within their structure, that is to say, an element which can belong to the structure of only one entity at a given time. Various versions of UE may be formulated by specifying the characteristics of the other elements of the structure, the characteristics of the unique element, or the relation between the structural elements. UE accounts are connected with substratum theories of individual objects (e.g. Allaire 1963) as well as with theories in which objects are individuated by non-relational locations (e.g. Quinton 1973). By combining UE with the criterion (*CON*) for constructing characteristics, we can put forward a hypothesis that CSSCs connected with Minimal Visual Objects differ from CSSCs shared by tokens of Maximal Non-object Representations by containing a description of a unique element.

The second account appeals to a "specific connection structure" (SC). According to this notion, elements that constitute objects are connected in a special way — for example, by a certain relation that binds them together. Obviously, it is possible to obtain different versions of SC by specifying the characteristics of elements which are combined and by describing more closely the tie that holds them together. SC accounts can mainly be found within the bundle theories of individual objects, both in their trope-based (e.g. Maurin 2002) and universal-based versions (e.g. O'Leary-Hawthorne, Carer 1998). By combining (*CON*) with SC, it may be proposed that the description of a specific connection differentiates CSSCs shared by tokens of Minimal Visual Objects from CSSCs shared by tokens of Maximal Non-object Representations.

The third notion of object structure is based on the concept of exemplifying a general category and may be called "general category structure" (GC). On this view, entities are objects by virtue of being the realization of a general category — preferably a kind that is the referent of a sortal term (usually a countable noun). Again, different versions of GC can be formulated by specifying the characteristics of a general category, of other structural elements, of relations between them, or of the relation of "exemplifying". GC is strongly connected with substantial theories of individual object (e.g. Loux 1978). (*CON*) together with GC suggest that CSSCs shared by tokens of Maximal Non-object Representations lack the description of exemplifying a general category present within CSSCs shared by tokens of Minimal Visual Objects.

The three above notions of the structure of visual objects are closely connected with the problem of individuation. According to UE notions, the sameness of the unique element is the necessary and sufficient condition of the sameness of objects (Denkel 1991). By contrast, in bundle theories associated with SC notions, there is no element that plays this special function. Every object is identical to a specific combination of elements and substituting one element for another results in a creation of a new object (Benovsky 2008). Finally, proponents of substantial theories connected with GC claim that the general category exemplified by an object determines its conditions of sameness, for example by designating essential and contingent features (Wiggins 1980).

All these views on individuation have some undesirable consequences when applied to visual objects. It is commonly claimed in scientific models of vision that the perceptual system can represent objects as being the same despite changes in their features and localization (Scholl 2007). Accounts adopting SC have difficulties in accommodating this observation since, according to them, every change produces a new object. What is more, if the unique element postulated in the UE notions is a location, then a similar problem arises, since visual objects can be represented as being the same even if they move through the visual field.

The possibility of persistence through change can be explained if the unique element is a "bare particular" that cannot be identified with any feature or location. However, such a solution comes at a price: it introduces elements which lack any obvious visual characteristics. Similarly, persistence of visual objects is compatible with GC notions. However, this has a consequence that all visual objects are exemplifications of some general categories. Such an assumption is quite implausible in the context of visual perception where processes of categorization occur at the highest stages of the perceptual process and visual objects may be represented by quite early perceptual mechanisms (Pylyshyn 2007). In what follows, I argue that the relational account of visual object construction, which is implicitly assumed by the models of figure/ground discrimination, offers a better explanation of the phenomenon of visual persistence.

The works by Clark and Raftopoulos contain descriptions of the structure of visual objects which bear a significant similarity to UE and SC. According to Clark, visual features have the characteristics of universals (Clark 2004: 453), in the sense that the same feature is realizable more than once, at a given time, in a single visual field, while every location always has a single realization. The feature-placing mechanism binds features with locations (Clark 2004: 453), while the attentional binding determines which features have a common location (Clark 2004: 446). As a result of these processes, co-localized feature-clusters are constructed. These more complex representations make it possible to distinguish between different instantiations of the same feature and to recognize which features are connected with the same location.

As opposed to Clark, Raftopoulos (2009: 91, Raftopoulos, Müller 2006b: 198) postulates that the first stage of visual object construction is governed by the preattentive segmentation and grouping processes. These mechanisms connect pre-object visual representation forming more complex wholes and organize them into visual regions. However, as mentioned above, some works by Raftopoulos (e.g. 2009: 7-15) contain statements that allow for an alternative interpretation, which is compatible with a relational view on visual object construction described further in the article.

Both Clark (2004: 455-456) and Raftopoulos (2009: 91, Raftopoulos, Müller 2006b: 198) claim that at the next stage of the perceptual process visual indices are attached to some of the representations constructed by features–location binding or by segmentation and grouping processes. The idea of visual indices is inspired by Pylyshyn's works (Pylyshyn 2001), who claims that a visual system is able to pick out several (4 to 6) objects from a scene and track them, preserving their identity despite movement or changes in their features.[2] Visual indices also serve as gates to the further stages of perceptual processing: indexed objects attract attention, giving access to visual awareness and facilitating the representation of their higher-level features, especially shape. Subsequently, the shape of an object can be compared with stored representations of various categories of objects, which leads to identifying an object as a member of a certain kind.

By using the above ideas, we can put forward three different hypotheses about the construction of visual objects:

(*H1*) Visual objects are formed by mechanisms that create localized feature-clusters.

(*H1*) implies that tokens of Minimal Visual Objects do not have a description of indices within their CSSCs. Here, the constructing characteristics are described by the combination of UE and SC notions — CSSCs of tokens of Minimal Visual Objects differ from CSSCs of tokens of Maximal Non-object Representations by including the description of a unique element (common location of different features) and of a specific connection between elements (attentional binding).

(*H2*) Visual objects are constructed by the preattentive segmentation and grouping processes.

(*H2*) also assumes that tokens of Minimal Visual Objects are not indexed. In this case, the constructing characteristics seem to be consistent with SC: tokens of Maximal Non-object Representations are transformed into tokens of Minimal Visual Objects when their elements are grouped in an appropriate way.

(*H3*) Visual objects are constructed by attaching visual indices.

According to (*H3*), tokens of Maximal Non-object Representations are transformed into tokens of Minimal Visual Objects by attaching visual indices. Visual indices are clearly unique elements: one index can be attached to only one visual object at a given time. Accordingly, the constructing characteristics in the case of (*H3*) will constitute an example of construction connected with UE.

In what follows, I test if the above hypotheses and ideas connected with metaphysical accounts of object structure are consistent with the constructing characteristics proposed by scientific models of figure/ground discrimination, and I argue that scientific models assume a different, more adequate relational notion.

2. VISUAL OBJECTS IN THE PERCEPTUAL PROCESS

In the previous section, I have presented some conceptual assumptions about the notion of visual object and alternative proposals of possible constructing characteris-

[2] The function of visual indices can also be served by "object files" (Kahneman et al. 1992).

tics. All these claims specify how I intend to analyze the problem of visual object construction. However, in order to conduct such investigations, one also needs to decide which scientific models are relevant to the issue. In the following paragraphs, I briefly describe the initial phases of the perceptual process so as to identify the stage at which visual objects are formed. Then I present scientific models connected with that stage, on the assumption that they may entail interesting statements about constructing characteristics of visual objects.

2.1. Stages of low-level vision

In presenting the stages of the visual process, I rely on the pattern of visual processing proposed by Rock and Palmer (1994b), whose general features are widely shared in the psychological literature. Rock and Palmer distinguish four phases of early visual information processing: (1) edge detection, (2) region formation, (3) figure/ground distinction, and (4) grouping/parsing.

During the edge detection stage, the regional borders of different light reflectance are represented as visual edges because of the neuron activity on the V1 level of the visual cortex (Hubel, Wiesel 1962). The low-level cortical structure is retinotopic: each neuron is activated by a stimulus coming from a certain region of the retina and the whole net of neurons map the retinal topography (Hubel 1995). Thus the edges can be localized. In the next step, local edges are connected according to Gestalt-like principles of grouping, especially by proximity (Rensink, Enns 1995, Palmer et al. 1996). As a result of the edge detection stage, edge maps are formed which present border nets on different scales and, after integrating information from both eyes, on different depths (Palmer, Rock 1994a). It is commonly accepted that this process happens unconsciously, preattentively, and in parallel to the whole visual field. Simple edges and their connections probably can be regarded as pre-object individuals, yet it is unlikely that they are visual objects. Still, in order to evaluate whether the type of representation that contains edges as tokens satisfies (*VO*), we have to investigate how edges are processed in the subsequent stages of the perceptual process.

At the region formation stage, information about edges and the features present on either side of them is used to divide the visual field into qualitatively homogenous regions. Various mechanisms of region formation have been proposed in the cognitive sciences, including: contour connecting and pixel classifying (Leung, Malik 1998, Bruce et al. 2000, Boykov, Jolly 2001), texture grouping based on local differences of their elements (Malik, Perora 1990, Nothdurft 1992), or interactions between mechanisms that connect edges and those which “fill in” closed borders with surface features (Grossenberg 1997). This leads to creating maps of localized features, on different scales and depth planes, surrounded by edges (Grossenberg 1994). It is controversial to what extent attention is needed to perform region formation, and whether attention is necessary for the process itself or only for awareness of the results (e.g. Ben-Av et al. 1992, Mack et al. 1992). According to Rock and Palmer (1994b), uniformly connected regions — closed by edges and filled by homogenous features — constitute object-candidates. According to other research, uniform connectedness is not necessary, and proximity between edges designating a region is sufficient for forming an object-candidate (Hon et al. 1999, Kimchi 2000). The term "object-candidate" means that some of the formed regions are later distinguished as figures, while others do not achieve that status and thus constitute ground.

Regions may be plausibly interpreted as representations resulting from Clark's (2004) feature-placing mechanism and attentional binding, as well as from the grouping and segmentation processes described by Raftopoulos (2009). Those representations are examples of Clark's feature-clusters, since they are composed of different kinds of features (e.g. colour, orientation, length) connected by their simultaneous presence at the same location. They can also be described in accordance with Raftopoulos' claims, as being constructed by grouping similar representations into larger wholes and by separating distinct ones.

Further region status is determined during figure/ground discrimination. Some regions become figures, while others are treated as the ground extending behind them. The regions distinguished as figures are then, i.e. in the subsequent grouping/parsing phase, grouped together as described by the Gestalt principles (e.g. by proximity or similarity — Kubovy et al. 1998, Elder, Goldberg 2002, Palmer, Beck 2007) or divided into parts (e.g. Hoffman, Richards 1984). Like in the case of the region formation stage, views differ as to the role of attention in making the figure/ground distinction: some researchers propose that without attention this process cannot be conducted (Mack et al. 1992), while others believe that attention is only needed for the conscious access to the distinction's results (Driver et al. 1992, Moore, Egeth 1997); one can also hold that only certain features, such as shape or precise localization, cannot be obtained without attention (Rock et al. 1992).

In the light of cognitive sciences, it seems that visual objects are constructed during figure/ground discrimination: in this phase of visual process certain representations are distinguished as figures from other non-object structures. The representation types present at other stages have characteristics which prohibit them from serving this function.

The structure of regions constructed during the region formation stage is not sufficient to constitute visual objects since some regions are interpreted as ground. Therefore, the Visual Region, as a type of visual representation, does not fulfil (*VO*): not all of its tokens are transformed into tokens of Developed Visual Representation, even under good processing conditions. The same conclusion applies to more primitive representation types, whose tokens are edges formed during the edge detection stage. They are not Visual Objects because some of their tokens are processed into regions which then form ground and are not transformed into tokens of Developed

Visual Representation.[3] The fourth stage — grouping and parsing — develops the structure of already constructed visual objects, so it contains representation types that are not Minimal Visual Objects.

Only during the stage of figure/ground discrimination do visual representations cross the border dividing the Maximal Non-object Representation and Minimal Visual Object. It seems that some tokens of the Visual Region type gain additional structural characteristics by virtue of which they will be further processed, eventually becoming tokens of the Developed Visual Representation. These regions, extended by an additional characteristic, seem to be tokens of the Minimal Visual Object, and their additional characteristic constitutes the constructing characteristic envisaged by (*CON*). The subsequent processing of these minimal visual objects includes attaching visual indices, grouping minimal objects into more complex ones, attracting attention, and categorizing. All stages after figure/ground discrimination are directed towards the creation of tokens of Developed Visual Representation: non-regressive processing at those stages will not produce tokens of any other type of visual representation.

In the next subsection, I describe the figure/ground discrimination stage by presenting scientific models which contain a detailed description of constructing characteristics that differentiates tokens of Minimal Visual Object from tokens of the Maximal Non-object Representation.

2.2. The figure/ground distinction

Of course, the distinction between figure and ground is easy to grasp from a phenomenological point of view (Palmer, Rock 1994b). Regions interpreted as figures are perceived, inter alia, as being closer, more salient, and borders between regions seem to belong to the figure region. By contrast, a ground-region is perceived as unconstrained by borders and extending behind a figure.

The scientific models of figure/ground distinction usually implement the idea of competition between regions (Pomerantz, Kubovy 1986, Palmer, Rock 1994a,b). It is often the case that one region gains the figure status, while the other is interpreted as ground. Which competing region will become a figure is determined by their geometrical features inferred from stimuli (so-called bottom-up factors) as well as by the higher-level perceptual mechanism connected with attention and categorization (top-down factors). The same intuition is expressed in some works by Raftopoulos (e.g. 2009: 7-15).

[3] It is not to say that representations of edge-structures — such as circles or squares — which we can see within our standard visual awareness are not visual objects. However, they are different representations than discontinuities formed at the first stages of the perceptual process. The representations of edge-structures that we know from our conscious vision are, in fact, quite thin, visual regions, and some of them may be distinguished as visual objects.

The first way to describe the feature/ground discrimination mechanism is to specify a set of rules characterizing relations that determine the competition outcome. For example, a region which is smaller, more symmetrical, more convex, has more closed contours, and/or lies closer to the observer will usually be interpreted as a figure. These bottom-up factors are supported by top-down mechanisms. In particular, having a shape that is more recognizable than the shape of another region and being under the focus of attention are factors which lead to figure status (Peterson et al. 1991, Peterson, Gibson 1993, 1994).

The second approach is to build a neural model of competitive figure/ground discrimination. Neurological findings on the behaviour of cells in the visual cortex have inspired such models. It is widely recognized that, even in the low-level cortical layers, neurons display different activities when they encode a figure fragment than when they encode a ground fragment (Lamme 1995, Albright, Stoner 2002). For example, so-called border ownership cells encode the information about the competing region to which a particular edge belongs (Zhou et al. 2000, Qiu, von der Heydt 2005). Importantly, whether a neuron displays the figure-activity or the ground-activity is not determined solely by the local feature of the stimuli received by that cell but also by the visual field's global content (Lee et al. 1998). Such findings are compatible with the idea of the competitive nature of figure/ground discrimination. Apparently, this process is not based on internal features of a region but relies instead on a relation between features of competing regions.

The general structure of the different neural figure/ground models consists of a few neural layers (Kienker et al. 1986, Vecera, O'Reilly 1998, Vecera 2000, Craft et al. 2007). Usually, the first layer is composed of "edge units". A few competing edge units are associated with every location in the visual field, and each of them encodes an alternative interpretation of border ownership. For example, if an edge is present in a location, then one edge-unit encodes the interpretation "the edge belongs to the figure on the left", while the other encodes the opposite interpretation "the edge belongs to the figure on the right". The choice between alternative edge units is made on the basis of bottom-up cues, similar to those proposed within the "set of relational rules" approach described earlier. For example, the edge is recognized as belonging to a more closed, smaller, more convex, and/or closer region. Some of the neural models assign object status to regions which have more closed contours than their competitors (e.g. Qiu et al. 2007).

The next neural layer consists of "figure units" that encode the presence of a figure fragment at a given location. Every edge unit is linked by an excitatory connection to a figure unit, and inhibitory connections hold between figure units connected to competing edge units. Competition-winning edge units facilitate the corresponding figure units, and by doing so they also inhibit the activation of other figure units. The multiple active figure units encode the region recognized as a figure.

In addition, neural competition models usually include top-down factors capable of resolving ambiguous cases. Some of these models (e.g. Vecera, O'Reilly 1998, Vecera 2000) postulate a categorization mechanism that recognizes a region as an exemplar of a certain general category of objects. According to that proposal, a pattern of active figure units determines the shape of a region which is then matched with higher-level representations of known object categories. The higher matching degree makes obtaining figure status more probable.

Other neural competition models (Kienker et al. 1986, Craft et al. 2007, Qiu et al. 2007) include a kind of attentional mechanism. Attention facilitates the activity of figure units encoding a certain region, leading to distinguishing it as a figure. Furthermore, moving the attention focus facilitates switching between the alternative figure/ground region divisions. Of course, these two proposals of top-down mechanisms should be treated as complementary rather than exclusive.

In the next section, I consider what constructing structural characteristics, understood in accordance with (*CON*), are suggested by the above models. Subsequently, it will be tested whether those characteristics are consistent with the "unique element", "specific connection", and "general category" approaches to visual object construction.

3. CONSTRUCTING CHARACTERISTICS IN SCIENTIFIC MODELS

3.1. The relational notion

The bottom-up factors which determine the figure/ground discrimination — proposed by neural (e.g. Vecera 2000) and "set of rules" models (e.g. Palmer, Rock 1994b) — are clearly relational. To become a figure, a visual region should be recognized as standing in certain relations to other visual regions. For example, a token of Visual Region becomes a token of Minimal Visual Object when it is recognized as more convex and/or as having more closed shape than its competitors.

By contrast, the relational nature of top-down factors is much less obvious. The first of the top-down mechanisms postulated in figure/ground discrimination models (e.g. Peterson, Gibson 1994, Vecera, O'Reilly 1998) is responsible for identifying visual regions as representing exemplars of general kinds, by matching their shape with representations stored in the memory. Having a shape that matches a category model is therefore a constructing characteristic which distinguishes tokens of Minimal Visual Objects from tokens of the Visual Region. However, many findings suggest that object construction achieved by such a matching also relies on relations. This can be illustrated by the well-known "Rubin vase" pattern.[4]

All regions within that pattern are, to some similar extent, exemplifications of general categories of objects: the central region resembles a vase and two sides of the black region are face-shaped. If the construction by matching a shape with a general

[4] See: http://designofthetimes.altervista.org/wp-content/uploads/2013/11/Rubin-vase-Edgar-John-Rubin-Synsoplevede-1915.jpg

kind model were non-relational, then all regions would be simultaneously recognized as figures, since they partially fall under a general category of objects. Yet people have difficulties in seeing all regions as objects and switch between seeing two faces or a vase.

It seems that the non-relational feature of being recognizable is not sufficient for achieving the visual object status: in addition, standing in a proper relation, for example of "being more recognizable", is needed. This claim can be demonstrated by means of a modified Rubin vase pattern, in which some of the regions match the general category to a higher degree.[5] In such a situation, a visual region that stands in a relation of being more recognizable than its neighbours is more likely to be recognized as a visual object.

Another top-down mechanism is connected with the focus of visual attention (e.g. Qiu et al. 2007). Being within the focus of attention allows a region to become a visual object, and by doing so the focus serves as a constructing characteristic. Judging whether the attentional figure/ground discrimination mechanism is relational — i.e. if a region becomes an object not when it is merely attended to but when, for example, it gets more attention than other regions — is a difficult task due to the variety of different models of visual attention offered in the psychological literature. According to some basic distinctions, attention could be:

(1) either unifocal or multifocal — either focused always on one part of the visual field or possessing an ability to be present in several disconnected parts (e.g. Cave and Bichot 1999, Müller et al. 2003),

(2) either uniformly distributed or gradient — either everything under the focus of attention receives the same amount of attentional resources or the attention strength decreases in a wave-like fashion from some point (or points) of the highest focus towards less attended-to areas (e.g. LaBerge, Brown 1989, Kramer, Jacobson 1991),

(3) either space-based or object-based — either directed upon a fragment of visual space or directed upon a visual representation (e.g. Scholl 2001, Sato, Blanco 2004).

Accordingly, combining these three pairs of features allows one to obtain eight alternative visual attention models. The relational interpretation is supported by all gradient models, regardless of whether they assume unifocality or multifocality, or whether they characterize attention as space- or object-based. In gradient models, the neighbouring visual regions may receive a different amount of attentional resources; usually a region in the centre of attentional focus is more attended to than its neighbours. In this article, I assume, in line with some important psychological works (e.g. Andersen, Kramer 1993, Dori, Henik 2006), that visual attention may be plausibly interpreted in terms of gradient models.

---

[5] See: http://www.moillusions.com/fats-rubin-vase-reveals-twin-portaits/

It follows that visual regions — clusters of visual features realized in a location designated by some edges — become visual objects when they are recognized as standing in proper relations, determined by both bottom-up and top-down factors, to other visual regions. Using the terminology associated with the general understanding of constructing characteristics presented in (*CON*), the CSSC shared by tokens of Minimal Visual Object (figure) is the expansion of the CSSC shared by tokens of Maximal Non-object Representation (visual region). The difference between these two sets consists in characteristics which describe standing in certain relations.

The CSSC of Visual Region's tokens consists of: a description of borders, a description of surface features realized within borders, and a description of localization. The CSSC shared by Minimal Visual Object's tokens additionally contains a description of relations, because in order to be constructed as visual objects they need a specific relational structure. These additional relational characteristics are constructing characteristics of visual objects, and I refer to the account invoking them as the "relational notion of visual object construction".

Three types of visual representations are involved here. The first one is the Visual Region, which plays the role of the Maximal Non-object Representation. Visual Region's tokens lack relational characteristics: at this stage relations between regions are not recognized by the visual system. When the relational characteristics of regions are recognized, some of them become tokens of Minimal Visual Object by virtue of their position in a certain relational network. Other regions, whose relations are less favourable, become tokens of non-object type of visual representation that contains elements of ground.[6]

3.2. The relational notion and other accounts

The above considerations show that the scientific models of figure/ground distinction which propose explanations in terms of competition between visual regions, as well as works in philosophy of perception which refer to these models (Raftopoulos 2009), assume a relational notion of visual object construction. This relational notion (REL) is significantly different from and cannot be easily reduced to accounts of construction associated with the most popular philosophical ideas concerning the structure of objects: the unique element structure (UE), the specific connection structure (SC), or the general category structure (GC).

[6] Of course, regions that constitute ground usually have some features that we intuitively associate with objects (like spatial coherence). However, their structure is different from the structure of minimal visual objects, and they are not determined to become tokens of Developed Visual Representation: even under good processing conditions it may be the case that they are not developed further. Hence I treat regions forming ground as tokens of a type of visual representation different from Minimal Visual Object.

REL is distinct from UE because relevant relations, such as "is more symmetrical than" or "is more attended to than", are not unique elements but are repeatable (that is, they can be simultaneously realized many times within a single visual field). The REL is also different from SC because object construction does not happen by virtue of relations between elements constituting a single region but due to relations between different regions. Finally, REL is also inconsistent with GC, since it is possible for a visual object to be formed without being immediately recognized as representing an instance of a general category. In addition, matching a shape with a general category seems to function as a kind of top-down relational factor.

From the above perspective, we can evaluate the accounts of visual object construction which can be found in works invoking mechanisms of feature-binding and region segmentation, together with hypotheses (*H1*)-(*H3*). According to these accounts, the constructing characteristics of visual objects are examples of UE or SC. However, the models of the figure/ground distinction suggest a different, relational notion of object construction, which is inconsistent with both UE and SC.

Furthermore, according to (*VO*), the hypotheses (*H1*) (construction by feature–place binding) and (*H2*) (construction by preattentive segmentation and grouping) seem to apply the label "visual object" too early. Every visual region is constructed by the mechanisms of binding, segmentation, or grouping, but not every visual region is transformed into a token of Developed Visual Representation, even under good processing conditions. The notions of object construction described by (*H1*) and (*H2*) do not explain why only select regions are processed in that way.

On the other hand, (*H3*) (construction by visual indices) attach the "visual object" label too late and creates an explanatory gap. If attaching an index is sufficient for creating a visual object from a pre-object visual region, then why can't the indices be attached to whichever visual region, allowing any of them to be transformed into a visual object? It seems that only some regions have characteristics that attract visual indices. From the perspective of (*CON*), attaching a visual index does not create tokens of Minimal Visual Object but operates on already formed minimal visual objects that were constructed from visual regions by virtue of their position in a proper relational network.

In addition, I stated earlier that UE, SC, and GC lead to certain problems with explaining the diachronic sameness of visual objects. REL, suggested by the models of figure/ground discrimination, is better suited for this task. If the structure of every visual object contains relations to other visual entities, then also criterions of diachronic sameness can be specified in terms of relations between visual objects, for example, some symmetric but intransitive similarity-like relations. One plausible connection is the relation of spatiotemporal continuity, which is often characterized as determining the sameness in the context of vision (Scholl 2007).

Such relational view of visual objects' sameness avoids the main problems related to UE, SC, and GC. Firstly, in contrast to GC, it does not assume that visual objects have to be represented as exemplars of general categories to be able to persist through change. The models of figure/ground discrimination postulate low-level mechanisms that resolve the competition between regions and operate before the categorization can take place. Secondly, there is no need for postulating a primitive non-qualitative particular element or for describing conditions of sameness with respect to the sameness of locations, which allows to avoid problems faced by UE. Of course, if spatiotemporal continuity is chosen as a relation determining the diachronic sameness of visual objects, then location plays a significant role, but sameness of location is not necessary for sameness of visual objects — only a weaker relation connected with certain spatial proximity is needed. Finally, in standard SC notions, related to the bundle theories of objects, sameness of objects is guaranteed only by sameness of all elements of a bundle, which makes the persistence through change impossible. The conditions of diachronic sameness allowed by REL simply sidestep this problem by stating that what determines sameness is not structural identity but some weaker relation between visual objects represented at the subsequent moments.

The general relational notion may be specified in many alternative ways. In the following paragraphs, I sketch a relational theory of visual object construction that accommodates the basic phenomena connected with figure/ground discrimination.

3.3. A theory of relational construction of visual objects

The first idea about the relations by virtue of which a visual region is transformed into a visual object comes from the notion of competition between regions. A region wins the competition when it is recognized as standing in appropriate relations to other visual regions, such as being more symmetrical, more convex or having more coherent borders. I call such relations, which are based on bottom-up factors constituted by geometrical features, the "bottom-up relations". The "worst" region — with respect to bottom-up relations — among all regions within a particular visual field is treated as ground and not as a visual object. What is "better" or "worse" in given cases is a matter of empirical investigations, which discover how certain differences between regions influence their chances during figure/ground discrimination.

Secondly, according to the observations made by Pylyshyn (2001), which inspired his notion of visual indices, it seems that a visual system is able to simultaneously form several (4 to 6) visual objects, which can be further processed, by attaching visual indices. According to this idea, a visual region, to become a token of Minimal Visual Object, needs to be among several best regions (with respect to bottom-up relations).

The above propositions need further modification if it is acknowledged that some visual regions may be mutually exclusive (as in the "Rubin vase" pattern). Functionally, the mutual exclusion (ME) between visual regions can be defined in the following simple way:

(ME) Two visual regions are mutually exclusive iff they cannot simultaneously be visual objects.

Cases of mutually exclusive visual regions lead to the conclusion that regions recognized as visual objects have to be better, with respect to bottom-up relations, than regions which are excluded by them.

Of course, visual object construction is not determined by bottom-up factors alone. A very important role is also played by top-down relations, especially those connected with visual attention. With respect to such relations, visual regions can be "less attended to" or "more attended to" than other regions. However, in some cases, attention may not differentiate among regions, allowing them to be equally attended to or not to be attended to at all. By adding this condition, we can formulate the theory of preattentive relational construction of visual objects:

(*PREL*) A visual region is transformed into a visual object if it is recognized that:

(1) it is not the worst of the formed regions with respect to bottom-up relations, and
(2) it is among the best N (number of visual indices) regions with respect to bottom-up relations, and
(3) it is better, with respect to bottom-up relations, than regions excluded by it, and
(4) attention does not differentiate between formed regions (they are not attended to at all or are equal with respect to top-down attentional relations).

Despite the fact that condition (4) refers to attention, (*PREL*) characterizes the preattentive construction of visual objects. If attention does not distinguish any region, then competition between regions is not resolved by relations connected with being attended to. To the contrary, in such a situation the figure-status of a visual region is determined solely by means of lower-level, preattentive mechanisms. Of course, it is also obviously true if the visual regions in question are not attended to at all.

However, the process of the preattentive construction is not able to deal with all relevant cases. In particular, PREL will not work if:

(1) the number or equally good (with respect to bottom-up relations) visual regions exceeds the number of visual indices;

(2) two mutually exclusive visual regions are equally good (with respect to bottom-up relations);

(3) according to a cognitive goal, such as searching for a camouflaged object, the visual system needs to recognize as an object a poor visual region (with respect to bottom-up relations).

Visual attention is able to resolve the above problems by (a) focusing only on some of the visual regions that are equal with respect to bottom-up relations, (b) focusing only on one of the mutually exclusive regions, and (c) focusing on the region which is worse, with respect to the bottom-up relations, than other regions. In all such cases, visual regions are transformed into visual objects by virtue of standing in top-down relations of being more attended to than others. What is more, the attentional top-down relations can override the bottom-up factors, so a highly attended-to region may be recognized as an object even if it is placed in an unfavourable bottom-up relational structure. By taking these remarks into consideration, we can put forward the theory of attentional relational construction of visual objects:

(*AREL*) A visual region is transformed into a visual object if it is recognized that:

(1) it is among the N (number of visual indices) regions that are equal with respect to top-down attentional relations, and those regions are more attended to than all other formed regions, and
(2) it is more attended to than any of the regions excluded by it.

(*PREL*) and (*AREL*) describe two sets of constructing characteristics that differentiate tokens of the Maximal Non-object Representation from tokens of the Minimal Visual Object. When a visual region is recognized as placed in a favourable bottom-up relational structure, and attention does not discriminate between regions, then it is transformed into a visual object preattentively — described by (*PREL*). If a region is recognized as standing in proper top-down attentional relations — described by (*AREL*) — then it is attentionally transformed into a visual object, even if its bottom-up relations are rather poor.

If a visual region satisfies conditions characterized by (*PREL*) or (*AREL*), then it also satisfies (*VO*) and is a minimal visual object. Such a region will then gain a visual index; its more complex features (especially its shape) will be processed; it can be grouped with other simple objects to form a more complex one; and finally, it will be categorized so as to become a developed visual representation. The only cases in which a region that possesses PREL or AREL characteristics would not reach developed visual representation status are those connected with a lack of time or cognitive capacities, or those caused by a regressive process that strips it of relevant relations and thereby turns it into an ordinary visual region, whose relational characteristics are not recognized.

It should be noted that on this account there are two distinct expansions of the set of structural characteristics of a visual region, and so two different types of representations that are Minimal Visual Objects. However, preattentively constructed objects usually attract attention and also gain favourable top-down relations.

4. CONCLUSIONS

I have argued that the adequate account of visual object construction can be provided by a relational notion that is assumed in the scientific models of figure/ground discrimination. According to this relational notion, a visual region is transformed into a visual object if and only if it is recognized as standing in certain relations to other regions. I have identified these relations as constructing characteristics of visual objects, since their presence distinguishes the CSSC shared by tokens of the Minimal Visual Object from the CSSC shared by tokens of the Visual Region, which are maximal non-object representations. The relational notion is related to a general idea according to which regions are distinguished as objects by means of comparing their features with those of other regions. More specifically, I have proposed that there are two types of relational constructing characteristics: one associated with the preattentive and another with the attentive mode of object formation.

Apart from explicating the relational notion of visual object construction, I have shown that the most popular notions of object structure proposed so far in analytic metaphysics, invoking unique elements, specific connections, or exemplification of general categories, fail to play an important role in describing the constructing characteristics of visual objects. In addition, the relational notion avoids problems with characterizing the diachronic sameness of visual objects, which are typical of other accounts.

Furthermore, the analysis suggests that accounts which rely only on mechanisms of feature-binding, regions segmentation, and attaching visual indices are insufficient to adequately characterize the transition from maximal non-object representations to minimal visual objects and so need adjustments to accommodate the important role of relational factors.

REFERENCES

Albright T. D., Stoner G. R. (2002), *Contextual Influences on Visual Processing*, „Annual Review of Neuroscience" 25(1), 339-379.

Allaire E. B. (1963), *Bare Particulars*, „Philosophical Studies" 14(1), 1-8.

Andersen G. J., Kramer A. F. (1993), *Limits of Focused Attention in Three-Dimensional Space*, „Perception and Psychophysics" 53(6), 658-667.

Ben-Av M. B., Sagi D., Braun J. (1992), *Visual Attention and Perceptual Grouping*, „Perception and Psychophysics" 52(3), 277-294.

Benovsky J. (2008), *The Bundle Theory and the Substratum Theory. Deadly Enemies or Twin Brothers?*, „Philosophical Studies" 141(2), 175-190.

Boykov Y., Jolly M.-P. (2001), *Interactive Graph Cuts for Optimal Boundary and Region Segmentation in N-D Images*, „Proceedings. Eighth IEEE International Conference on Computer Vision", ICCV 2001, vol. 1, 105-112.

Bruce J., Balch T., Veloso M. (2000), *Fast and Inexpensive Color Image Segmentation for Interactive Robots*, „Proceedings. 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems", IEEE, vol. 3, 2061-2066.

Cave K. R., Bichot N. P. (1999), *Visuospatial Attention. Beyond a Spotlight Model*, „Psychonomic Bulletin and Review" 6(2), 204-223.

Clark A. (2004), *Feature-Placing and Proto-objects*, „Philosophical Psychology" 17(4), 443-469.

Craft E., Schütze H. Niebur E., von der Heydt R. (2007), *A Neural Model of Figure–Ground Organization*, „Journal of Neurophysiology" 97(6), 4310-4326.

Denkel A. (1991), *Principia Individuationis*, „The Philosophical Quarterly" 41(103), 212-228.

Dori H., Henik A. (2006), *Indications for Two Attentional Gradients in Endogenous Visual-Spatial Attention*, „Visual Cognition" 13(2), 166-201.

Driver J., Baylis G. C., Rafal R. D. (1992), *Preserved Figure–Ground Segregation and Symmetry Perception in Visual Neglect*, „Nature" 360(6399), 73-75.

Elder J. H., Goldberg R. M. (2002), *Ecological Statistics of Gestalt Laws for the Perceptual Organization of Contours*, „Journal of Vision" 2(4), 324-353.

Grossenberg S. (1994), *3-D Vision on Figure–Ground Separation by Visual Cortex*, „Perception and Psychophysics" 55(1), 48-120.

Grossenberg S. (1997), *Cortical Dynamics of Three Dimensional Figure–Ground Perception of Two-Dimensional Pictures*, „Psychological Review" 104(3), 618-658.

Hoffman D. D., Richards W. A. (1984), *Parts of Recognition*, „Cognition" 18(1-3), 65-96.

Hon S., Humphreys W., Chen L. (1999), *Uniform Connectedness and Classical Gestalt Principles of Perceptual Grouping*, „Perception and Psychophysics" 61(4), 661-674.

Hubel D. H. (1995), *Eye, Brain, and Vision*, http://hubel.med.harvard.edu/book/bcontex.htm.

Hubel D. H., Wiesel T. N. (1962), *Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex*, „Journal of Physiology" 160(1), 106-154.

Kahneman D., Treisman A., Gibbs B. J. (1992), *The Reviewing of Object Files. Object Specific Integration of Information*, „Cognitive Psychology" 24(2), 175-219.

Kienker P. K., Sejnowski T. J., Hinton G. E., Schumacher E. (1986), *Separating Figure from Ground with a Parallel Network*, „Perception" 15(2), 197-216.

Kimchi R. (2000), *The Perceptual Organization of Visual Objects. A Microgenetic Analysis*, „Vision Research" 40(10-12), 1333-1347.

Kramer A. F., Jacobson A. (1991), *Perceptual Organization and Focused Attention. The Role of Objects and Proximity in Visual Processing*, „Perception and Psychophysics" 50(3), 267-284.

Kubovy M., Holcombe A. O., Wagemans J. (1998), *On the Lawfulness of Grouping by Proximity*, „Cognitive Psychology" 35(1), 71-98.

LaBerge D., Brown V. (1989), *Theory of Attentional Operations in Shape Identification*, „Psychological Review" 96(1), 101-124.

Lamme V. A. F. (1995), *The Neurophysiology of Figure–Ground Segregation in Primary Visual Cortex*, „The Journal of Neuroscience" 15(2), 1606-1615.

Lee T. S., Mumford D., Romero R., Lamme V. A. F. (1998), *The Role of the Primary Visual Cortex in the Higher Level Vision*, „Vision Research" 38(15-16), 2429-2454.

Leung T., Malik J. (1998), *Contour Continuity in Region Based Image Segmentation* [in:] *Computer Vision — ECCV' 98*, *Vol. I, LNCS 1406*, H. Burkhardt, B. Neuman (eds.), Berlin: Springer, 514-559.

Loux M. (1978), *Substance and Attribute*, Dordrecht: D. Reidel.

Lowe E. J. (2006), *The Four-Category Metaphysics*, New York, NY: Oxford University Press.

Mack A., Tang B., Tuma R., Kahn S., Rock I. (1992), *Perceptual Organization and Attention*, „Cognitive Psychology" 24(4), 475-501.

Malik J., Perora P. (1990), *Preattentive Texture Discrimination with Early Vision Mechanism*, „Journal of the Optical Society of America" 7(5), 923-932.

Marr D. (1982), *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*, Cambridge, MA: MIT Press.

Martin C. B. (1980), *Substance Substantiated*, „Australasian Journal of Philosophy" 58(1), 3-10.

Maurin A.-S. (2002), *If Tropes*, Dordrecht: Kluwer.

Moore C. M., Egeth H. (1997), *Perception without Attention. Evidence of Grouping under Conditions of Inattention*, „Journal of Experimental Psychology: Human Perception and Performance" 23(2), 339-252.

Müller M. M., Malinowski P., Gruber T., Hillyard S. A. (2003), *Sustained Division of the Attentional Spotlight*, „Nature" 424(6946), 309-312.

Nothdurft H.-C. (1992), *Feature Analysis and the Role of Similarity in Preattentive Vision*, „Perception and Psychophysics" 52(4), 355-375.

O'Leary-Hawthorne J., Carer J. A. (1998), *A World of Universals*, „Philosophical Studies" 91(3), 205-219.

Palmer S. E. (1999), *Vision Science. Photons to Phenomenology*, Cambridge, MA: MIT Press.

Palmer S. E., Beck D. M. (2007), *The Repetition Discrimination Task. An Objective Method of Study Perceptual Grouping*, „Perception and Psychophysics" 69(1), 68-78.

Palmer S. E., Neff J., Black D. (1996), *Late Influences on Perceptual Grouping. Amodal Completion*, „Psychonomic Bulletin and Review" 3(1), 75-80.

Palmer S. E., Rock I. (1994a), *On the Nature and Order of Organizational Processing. A Replay to Peterson*, „Psychonomic Bulletin and Review" 1(4), 515-519.

Palmer S. E., Rock I. (1994b), *Rethinking Perceptual Organization. The Role of Uniform Connectedness*, „Psychonomic Bulletin and Review" 1(1), 29-55.

Peterson M. A., Gibson B. S. (1993), *Shape Recognition Inputs to Figure–Ground Organization in the Three-Dimensional Displays*, „Cognitive Psychology" 25(3), 383-429.

Peterson M. A., Gibson B. S. (1994), *Object Recognition Contributions to Figure–Ground Organization: Operations on Outlines and Subjective Contours*, „Perception and Psychophysics" 56(5), 551-564.

Peterson M. A., Harvey E. M., Weidenbacher H. J. (1991), *Shape Recognition Contributions to Figure Ground Reversal. Which Route Counts?*, „Journal of Experimental Psychology: Human Perception and Performance" 17(4), 1075-1089.

Pomerantz J. R., Kubovy M. (1986), *Theoretical Approaches to Perceptual Organization* [in:] *Handbook of Perception and Human Performance*, K. R. Boff, L. Kaufman, J. P. Thomas (eds.), New York, NY: Wiley, 1-46.

Pylyshyn Z. W. (2001), *Visual Indexes, Preconceptual Objects, and Situated Vision*, „Cognition" 80(1-2), 127-158.

Pylyshyn Z. W. (2007), *Things and Places*, Cambridge, MA: MIT Press.

Qiu F. T., Sugikara T., von der Heydt R. (2007), *Figure–Ground Mechanisms Provide Structure for Selective Attention*, „Nature Neuroscience" 10(11), 1492-1499.

Qiu F. T., von der Heydt R. (2005), *Figure and Ground in the Visual Cortex. V2 Combines Stereoscopic Cues with Gestalt Rules*, „Neuron" 47(1), 155-166.

Quinton A. (1973), *The Nature of Things*, London: Routledge and Kegan Paul.

Raftopoulos A. (2009), *Cognition and Perception. How Do Neural Science and Cognitive Psychology Inform Philosophy?*, Cambridge, MA: MIT Press.

Raftopoulos A., Müller V. C. (2006a), *Nonconceptual Demonstrative Reference*, „Philosophy and Phenomenological Research" 72(2), 251-285.

Raftopoulos A., Müller V. C. (2006b), *The Phenomenal Content of Experience*, „Mind and Language" 21(2), 187-219.

Rensink R. A., Enns J. T. (1995), *Preemption Effects in Visual Search. Evidence for Low-Level Grouping*, „Psychological Review" 102(1), 101-130.

Sato D., Blanco M. J. (2004), *Spatial Attention and Object Attention. A Comparison within a Single Task*, „Vision Research" 44(1), 69-81.

Scholl B. J. (2001), *Objects and Attention. The State of the Art*, „Cognition" 80(1-2), 1-46.

Scholl B. J. (2007), *Object Persistence in Philosophy and Psychology*, „Mind and Language", 22(5), 563-591.

Spelke E. S. (1990), *Principles of Object Perception*, „Cognitive Science" 14(1), 29-56.

Treisman A. (1999), *Solutions to the Binding Problem. Progress through Controversy and Convergence*, „Neuron" 24(1), 105-110.

Van Cleve J. (1985), *Three Versions of the Bundle Theory*, „Philosophical Studies" 47(1), 95-107.

Vecera S. P. (2000), *Towards a Biased Competition Account of Object-Based Segregation and Attention*, „Brain and Mind" 1, 353-384.

Vecera S. P., O'Reilly R. C. (1998), *Figure–Ground Organization and Object Recognition Processes. An Interactive Account*, „Journal of Experimental Philosophy: Human Perception and Performance" 24(2), 441-462.

Wiggins D. (1980), *Sameness and Substance*, Oxford: Blackwell.

Zhou H., Friedman H. S., von der Heydt R. (2000), *Coding of Border-Ownership in Monkey Visual Cortex*, „The Journal of Neuroscience" 20(17), 6594-6611.