Relational Construction of Visual Objects

Błażej Skrzypulec

One of the main functions of visual system is to construct representations of objects. These visual objects are formed by developing and modifying the structure of more primitive visual representations. In the paper, I consider the type of structural change that transforms visual regions, which may be treated as maximal non-object representations, into basic, low-level visual objects. In order to identify such structural development, theories of early vision, models of figure/ground distinction, and ontological notions of objects' structure are analyzed and discussed. A thesis is stated that external relations between visual representations are the main structural factor that determines the transition between non-object representations and visual objects. I argue that such an approach is more adequate than propositions according to which visual objects are distinguished by virtue of internal relations, possessing certain features, exemplifying a general category, or having a visual index attached. This conclusion is preceded by conceptual considerations by which I specify a criterion for distinguishing category of visual objects from other types of visual representations.

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

Within this context, one asks what concept of object is able to adequately describe the way in which visual objects are formed. Psychological and neuroscientific models often consider processes that allow the construction of visual objects (e.g., Treisman, 1999). Unfortunately, they usually do not make all their conceptual assumptions and consequences explicit, nor do they sufficiently explain which philosophical concept of object is used within their framework.

On the other hand, various conceptions of objects' structure have been proposed by philosophical theories of individual objects, formulated on the grounds of the analytic metaphysics. The philosophical debates have mainly considered conceptions offered by the substratum (e.g., Martin, 1980), bundle (e.g., van Cleve, 1985), and substantial (e.g., Lowe, 2006) theories. However, in discussions regarding those metaphysical conceptions, scientific theories concerning vision are rarely addressed and the proposed conceptions of objects' structure were not usually applied to visual objects.

Within this article, I will consider two examples of philosophical investigations about the structure of visual objects that can be found in works by Austen Clark (2004) and Athanassios Raftopoulos (2009). I address their conceptions by proposing how non-object visual representations are transformed into visual objects in the process of distinguishing figures from the ground. Despite accepting their general views concerning preattentive vision, I will argue that feature-placing mechanisms (proposed by Clark) as well as segmentation and grouping processes (proposed by Raftopoulos) are not sufficient 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).

I start with conceptual considerations about the understanding of key notions such as "visual object" and "constructing characteristics". Firstly, a criterion for distinguishing visual objects from other types of visual representations is proposed. Secondly, the notion of "constructing characteristics" is explicated. Then, I investigate in which stage of the perceptual process visual objects are formed. The contemporary scientific models connected with that stage are described, and their conceptual assumptions, concerning the construction of visual objects, are revealed. Finally, the coherence of the most popular contemporary philosophical notions of objects' structure and the assumptions of the scientific models considered is tested. As a result, I conclude that relational factors are the most relevant in the context of visual objects' construction.

I. Conceptual Considerations

a) Representation and Visual Objects

According to the constructivist paradigm of vision science (e.g., Palmer, 1999), a visual object is characterized as a representation. The perceptual system relies on the information received on the retina and constructs visual objects which represent the source of the incoming stimuli.

At this point, it should be noted that I do not use the term "object" in a logical sense—as anything to which predicates can be ascribed. In such case, everything presented in the visual field would constitute a visual object. Instead, I follow a common-sense notion that the label "object" is restricted to only some of the patterns. For instance, a black square on a white background is clearly an object, but there is no object formed by two tiny spots lying at opposite edges of a visual field.

If we want to more precisely grasp what a visual object is, then three elements connected with the notion of 'representation' should be distinguished:

1) Physical object—visual objects represent postulated 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 visual objects' structure represents the physical objects' structure quite accurately in ordinary conditions. In this article, I will not investigate further neither the structure of physical objects, nor the representational relationship that connects them with visual 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", but 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 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', and I 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, color, etc.). In such a situation, visual objects' structure consists of at least two distinct kinds of elements (i.e., location and features[1]) combined by some relation. This structure does not necessarily match the structure of the external, physical object, and quite clearly is different form the structure of the neural mechanisms that contributed to the creation of the visual object.

b) Which Representations are Visual Objects?

Visual objects are not the only 'inhabitants' of the visual field. There are also various pre-object representations—such as edges, blobs, regions constituting ground behind figures etc. (Marr, 1982; Palmer, Rock, 1994b). Due to this, a conceptual problem arises: What is the criterion for deciding whether a certain type of visual representation is a visual object? For example, Clark claims that a visual object is everything that a visual system treats as an object

[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.

(Clark, 2004, pp. 463). In fact, it is easy to recognize objects in our usual visual field, when we have conscious access to the final product of the visual system. It is much harder to decide what is recognized as an object within the earlier stages of the perceptual process, which we are usually not aware of.

The standard way to omit that problem is to make an abstraction from the features of fully developed visual objects. With such an approach, more primitive visual objects share some characteristics with fully developed ones: for example, they are spatially coherent (Spelke, 1990) or are able to persist through change (Raftopoulos, Müller, 2006a, pp. 254), though they still lack some of their features (in that they are not identified as members of a general category). Unfortunately, the choice of the relevant features will always be arbitrary to some degree and it is quite probable that all visual representations, including the most primitive, share some features with fully developed visual objects. In addition, in the standard approach what the visual system treats as objects is hardly investigated—rather it utilizes our intuitions about what it means to be an object.

In this paper I choose a different criterion. I do not start from judging whether a type of visual representation is a visual object based on features shared by its tokens, but rather by investigating the placement 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 a perceiver's environment as entities that are exemplars of general categories, have a definite shape (often, but not always, in a form of complex 3D structure), and are accessible within the standard visual awareness.

Different, more primitive visual representation types are related to the DVR by some process. The relation that is especially relevant within this paper's context may be called 'rigid connection':

(RC) A type of visual representation is rigidly connected to the developed visual representation 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, turning it into a more basic one. Regressive processes arise in cases 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 the DVR. Meaning, that, even under very good processing conditions, at least some of their tokens are not transformed into tokens of the developed visual representation.

Relying on *DVR* and *RC* the notion of 'visual object' can be characterized as:

(VO) A visual object is a type of visual representation that is identical with the developed visual representation or that is rigidly connected to the 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. It allows to identify what is treated as a visual object from the point of view of visual system 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 the DVR as its highest element. At the bottom of the hierarchy are the most primitive types of representations which are still rigidly connected to the developed visual representation—i.e. all tokens of those types are transformed into DVR's tokens under good processing conditions.

Later in the paper, I investigate, referring to the scientific models, at what stage of visual processing representations are created that belong to the types that are rigidly connected to the developed visual representation. To avoid confusion, a terminological convention is adopted, that in order to refer to types of visual representations capital letters are used (e.g., Visual Object), while their tokens are named by using letters in lowercase (e.g., visual object).

c) Constructing Characteristics

The previous paragraphs presented what I understand by the term 'visual object' and proposed the criterion for distinguishing visual objects from different types of visual representations. Relying on that, it can be now explained, what I mean by 'construction of visual object'.

The *VO* criterion decides whether representations are tokens of a Visual Object, by considering their place within the perceptual process. In the subsequent parts 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. What is more, 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 the notions of ‘minimal visual object’ and ‘maximal non-object representation’ may be introduced:

(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 that 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—that is still a Visual Object. It is worth noting, that there may be several types of representations 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 that 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 that 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 connected with a type that is a Minimal Visual Object. Similarly as in 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.

The visual objects’ 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 identify which structural characteristics may be identified as 'constructing' and which types of ontological elements they describe.

d) Objects' Structure in the Philosophical Conceptions

On the analytic metaphysical grounds, the three general ideas about objects' structure have become the most influential. The first of those three notions henceforth will be named 'unique element structure (UE)'. According to this notion, entities are objects only when they possess a unique element within their structure—that means an element which can belong to the structure of only one entity at a given time. Various UE versions may be formulated by specifying the characteristics of the other elements of objects' structure, the characteristics of the unique element, or the relation in which structural elements stand. The UE conceptions are connected with the substratum theories of individual objects (e.g., Allaire, 1963) as well as the conceptions in which objects are individuated by non-relational locations (e.g., Quinton, 1973). By combining the *CON* criterion for constructing characteristics together with UE, a hypothesis can be stated 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 conception describes 'specific connection structure (SC)'. Due 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 that are combined and by describing more closely the bind that keeps them. SC conceptions can mainly be found within the bundle theories of individual objects, both in their universals (e.g., O'Leary-Hawthorne, Carer, 1998) and tropes variants (e.g., Maurin, 2002). By combining the *CON* criterion together 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 objects' structure is based on the concept of exemplifying a general category and may be named 'general category structure (GC)'. Within that conception, 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). The *CON* criterion together with GC suggests 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 works by Clark and Raftopoulos contain descriptions of visual objects' structure which bear a significant similarity to UE and SC notions. According to Clark, visual features have the characteristics of universals (Clark, 2004, pp. 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, pp. 453); while the attentional binding identifies which features have a common location (Clark, 2004, pp. 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 postulates (Raftopoulos, Müller, 2006b, pp. 198; Raftopoulos, 2009, pp. 91) 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.

Both Clark (2004, pp. 455-456) and Raftopoulos (Raftopoulos, Müller 2006b, pp. 198; Raftopoulos 2009, pp. 91;) claim that in the next stage of the perceptual process the visual indices are attached to some of the representations constructed as a result of features-location binding or by segmentation and grouping processes. The idea of visual indices is inspired by Pylyshyn's works (Pylyshyn, 2001), in which he proposes 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, which give access to visual awareness and facilitate the representation of higher-level features, especially shape. Subsequently, an object's shape can be compared with stored representations of different categories of objects that leads to identifying an object as a member of a certain kind.

By using the notions proposed by Clark and Raftopoulos three different hypotheses about the construction of visual objects can be proposed:

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

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

H1 implies that tokens of Minimal Visual Objects do not have a description of indices within their CSSCs. In that case, the constructing characteristics are described by the combination of the UE and SC notions—CSSCs of Minimal Visual Objects' tokens differs from CSSCs of Maximal Non-Object Representations' tokens 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 context, the constructing characteristics seem to be consistent with the SC notion—tokens of Maximal Non-Object Representations are transformed into Minimal Visual Objects' tokens when their elements are grouped in some appropriate way.

H3) Visual objects are constructed by attaching the visual indices.

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

Later in this article, I test if the above hypotheses and ideas connected with metaphysical conceptions of objects' structure are consistent with the constructing characteristics proposed by scientific models of figure/ground distinction, and I argue that a different relational notion is more adequate.

II. Visual Objects in the Perceptual Process

In the previous section, I presented some conceptual assumptions about the notion of 'visual objects' and alternative propositions of possible constructing characteristics. All these claims specify how I intend to analyze the problem of a visual object's construction. However, to conduct such investigations, it is also needed to decide which scientific models are relevant for the issue. In the following paragraphs, I briefly describe the initial phases of the perceptual process in order to identify a stage in which visual objects are formed. Then, I present scientific models connected with that stage, assuming that they may imply interesting statements about visual objects' constructing characteristics.

a) 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 within 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 due to the neuron activity in 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. The simple edges and their connections probably can be regarded as pre-object individuals, but they are unlikely to be visual objects. However, to evaluate whether the type of representation that contains edges as tokens satisfies the *VO* criterion, we have to investigate how edges are processed in the subsequent stages of the perceptual process.

Within 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 by: 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 the interactions between mechanisms that connect edges and those which 'fill in' closed borders with surface features (Grossenberg, 1997). The result is the creation of localized features maps, 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), the uniformly connected regions—closed by edges and filled by homogenous features—constitute object-candidates. Due 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 rank 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, as they are composed of different kinds of features (e.g. color, orientation, length) connected by their simultaneous presence at the same location. They can be also 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 ground extending behind them. Those regions distinguished as figures are then, that is, 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). As in the region formation stage, views differ on 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); or they hold that only certain objects' features, like shape or precise localization, cannot be obtained without attention (Rock et al., 1992; Rensink, Enns, 1995).

According to the cognitive science literature, it seems that figure/ground discrimination is a visual process phase in which objects are constructed and during that phase certain representations are distinguished as figures from other non-object structures. The representations types present at the other stages have characteristics which prohibit them from serving that very function.

The structure of regions constructed during the region formation stage is not sufficient to constitute visual objects as some of the regions are interpreted as ground. Because of that the Visual Region, as a type of visual representation, does not fulfill the *VO* criterion—not all of its tokens are transformed into Developed Visual Representation's tokens, even in good processing conditions. Consequently, the same conclusion applies to more primitive representation types, whose tokens are edges formed during the edge detection stage. They are not Visual Objects due to the fact, that some of their tokens are processed into regions which then form ground and are not transformed into Developed Visual Representation's tokens. 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 processed further (as stated in the *VO* criterion); finally 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 constitute the constructing characteristic according to the *CON* criterion. 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 Developed Visual Representation's tokens—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 detail description of constructing characteristics that differs Minimal Visual Object's tokens from tokens of the Maximal Non-Object Representation.

b) The Figure/Ground Distinction

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

The scientific models of figure/ground distinction usually implement the competition between regions idea (Pomerantz, Kubovy, 1986; Palmer, Rock, 1994a; 1994b). Often one will gain the figure status, while the other will be interpreted as ground. Which competing region will become a figure is determined by their geometrical features inferred from stimuli (so-called 'bottom-up' factors) and also by the functioning of the higher-level perceptual mechanism connected with attention and categorization ('top-down' factors).

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, usually a region that is smaller, more symmetrical, and more convex, has more closed contours, and/or which lies closer to the observer will 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 behavior of cells in the visual cortex have inspired such models. It is widely recognized that, even in the low-level cortical layers, neurons present 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 to which competing region a particular edge belongs (Zhou et al., 2000; Qiu, von der Heydt, 2005). Importantly, whether a neuron presents the figure-activity or the ground-activity is not determined solely upon 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 coherent with the idea of the competitive nature of figure/ground discrimination. It seems that this process is not based on a region's internal features, but rather relies on a relation between competing regions' features.

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 the alternative interpretations 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'. Choice between the 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 those 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.

The competition neural models usually also include top-down factors able to resolve ambiguous cases. Within some of these models (e.g., Vecera, O'Reilly, 1998; Vecera, 2000), a categorization mechanism that recognizes a region as an exemplar of a certain general category of objects is postulated. 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. What is more, 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 as with the *CON* criterion, are suggested by the above models. Subsequently, it will be tested whether those characteristics are consistent with 'unique element', 'specific connection', and 'general category' approaches to the objects' construction.

III. Constructing Characteristics in the Scientific Models

a) The Relational Notion

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

On the contrary, the relational nature of top-down factors is much less obvious. The first of the top-down mechanisms postulated within figure/ground discrimination models (e.g., Peterson, Gibson, 1994; Vecera, O'Reilly, 1998) is responsible for identifying regions as exemplifications of general kinds, by matching their shape with representations stored in the memory. Exemplifying a kind is therefore a constructing characteristic which distinguishes tokens of Minimal Visual Objects from tokens of the Visual Region. However, many demonstrations suggest that the objects' construction by exemplifying a general kind also relies on relations. This can be illustrated by the well-known pattern presented in Figure 1.

Figure 1 An example of the classical Rubin vase pattern.
Source: http://www.positivehealth.com/article/holistic-psychotherapy/gestalt-therapy-and-humanistic-psychology (Litt, 2000)

All regions within that pattern are, to some similar degree, exemplifications of general categories of objects—the central region resembles a vase and two parts of the black region are face-shaped. If the construction by exemplifying a general kind was non-relational, then all regions would be simultaneously recognized as figures, as they partially fall under a general category of objects. On the contrary, people have difficulties in seeing all regions as objects and they switch between seeing two faces or a vase on the objectless ground.

It seems that a non-relational feature of being recognizable is not sufficient for achieving the visual object status, but, in addition, standing in a proper relation, for example of 'being more recognizable', is needed. That claim is demonstrated by means of a modified Rubin vase pattern (Figure 2), in which some of the regions match the general category to a higher degree. In such a situation, a region that stands in a relation of being more recognizable than its neighbors is more likely to be recognized as a visual object.

Figure 2 The modified Rubin vase pattern.
Source: http://www.moillusions.com/2008/09/fats-rubin-vase-reveals-twin-portaits.html (Vurdlak, 2008).

The second 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 it is, for example, attended to more than other regions—is a difficult task due to the variety of different models of visual attention offered within the psychological literature. According to some basic distinctions, attention could be:

(1) either unifocal or multifocal—either focused always in one part of the visual field or possessing an ability to be present in a few disconnected parts (e.g., Cave, 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).

Consequently, combining these three pairs of features easily allows one to obtain eight alternative visual attention models. The relational interpretation is supported by all gradient models. In these models, the neighboring visual regions may receive a different amount of attentional resources; usually a region in the center of attentional focus is attended to more than its neighbors. Within this article, I assume, in line with the important body of contemporary psychological works (e.g., Andersen, Kramer, 1993; Dori, Henik, 2003), that visual attention may be plausibly interpreted as proposed in the gradient models.

From the above considerations the inference can be drawn 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 connected with the general understanding of constructing characteristics presented earlier (the *CON* criterion); the CSSC shared by Minimal Visual Object's (figure) tokens is the expansion of the CSSC shared by tokens of Maximal Non-Object Representation (visual region). The difference between those two sets consists of characteristics describing 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 to be constructed as a visual object specific relational structure is needed. Those additional relational characteristics are constructing characteristics of visual objects and I refer to the conception relying on them as the 'relational notion of objects' construction'.

Three types of visual representations are involved here. The first one is the Visual Region that plays the role of the Maximal Non-Object Representation. Visual Region's tokens do not have relational characteristics—at that stage relations between regions are not recognized by the visual system. When the relational characteristics of regions are recognized, some of them become Minimal Visual Object's tokens, by virtue of standing in a certain relational network. Other regions, which relations are less favorable, become tokens of non-object type of visual representation that contains elements of ground.

b) The Relational Notion and Other Conceptions

The relational notion of visual objects' construction (REL) cannot be easily reduced to construction conceptions connected with the philosophical ideas concerning objects' structure: the unique element structure (UE), the specific connection structure (SC), or the general category structure (GC). REL is distinct from the UE because relevant relations, such as 'is more symmetrical than' or 'is attended to more 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 the SC due to the fact that objects' construction does not happen by virtue of relations between elements constituting a single region, but because of relations between different regions. Finally, the REL is also inconsistent with the GC idea, because it is possible for a visual object to be formed without being immediately recognized as an instance of a general category of objects. In addition, exemplifying a general category seems to function as a kind of top-down relational factor.

From the above perspectives, the conceptions of a visual object's construction that can be found in works by Clark and Raftopoulos, together with hypotheses H1, H2 and H3, can be evaluated. According to these notions, the constructing characteristics of visual objects are examples of the UE or the SC kind. However, the models of the figure/ground distinction suggest a different, relational notion of objects' construction, which is inconsistent with both UE and SC.[3]

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

On the other hand, the hypothesis H3 (construction by the visual indices) introduces the 'visual object' notion 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 the indices not be attached to whichever visual region, allowing any of them to be transformed into a visual object. On the contrary, it seems that only some regions have certain characteristics that attract visual indices. From the *CON* criterion perspective, 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 standing in a proper relational network.

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

c) *A Theory of Visual Objects' Relational Construction*

[3] However, it may be possible to reinterpret the notion proposed by Raftopoulos in a relational fashion, due to the fact that he also describes the development of visual representations as competition (Raftopoulos, 2009, pp. 7-15).

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 proper relations to other visual regions, such as when it is more symmetrical, more convex or has more coherent borders. Such relations, which are based on bottom-up factors constituted by geometrical features, I call the 'bottom-up relations'. The 'worst' region according to possessed bottom-up relations, among all within a particular visual field, is treated as ground and not as a visual object. What is 'better' or 'worse' in considered 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) that inspired his visual indices notion, 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 the several best regions according to bottom-up relations.

The above propositions need further modification if it is acknowledged that some visual regions may be mutually exclusive. 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.

There are at least two kinds of ME. The first of them is connected with the top-down categorization process. The probability of such a top-down ME is high when regions are perceived as equally distant from the observer, match general categories to a similar degree, and share a substantial portion of their borders (Figure 1).

The second kind occurs by virtue of bottom-up geometrical factors relevant for the visual completion phenomenon. Visual completions happen when one region is recognized as being completed behind the other region.[4] In the psychological literature, models of visual completion explain the occurrence of that phenomenon by recognizing certain features of borders discontinuities and ways in which edges could be bound behind the occluder (e.g., Kellman, Shipley, 1991; Tse, 1999).

The visual completion phenomenon may lead to ME when completion is symmetrical, that is when it is possible to interpret one region as occluded by another and vice versa. Figure

[4] For arguments that visual completion is an early process, occurring preattentively, see Moore et al. (1998).

3 demonstrates such a situation: The black cross may be perceived as an object on the white ground completed behind it, but it is also possible to see the white 'fan' as a figure on black ground.

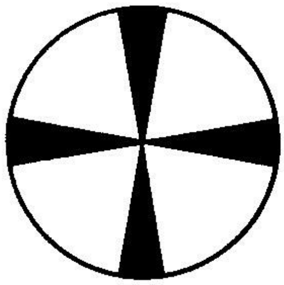

Figure 3 Symmetrical completion leads to competition.
Source: http://www.tau.ac.il/~tsurxx/FigureGround/Figure-ground+mp3New.html (Tsur, 1998).

The ME case inclusion leads to the conclusion that visual regions recognized as visual objects have to be better, according to bottom-up relations, than their mutually exclusive regions.

Of course, the visual objects' construction is not determined only by bottom-up factors. A very important role is also played by top-down relations, especially those connected with visual attention. According to such relations visual regions can be 'attended to less' or 'attended to more' than other regions. In some cases, attention may not differentiate among regions, allowing them to be attended to equally. By adding this condition the theory of preattentive relational construction of visual objects can be formulated:

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

(1) it is not the worst of the presented regions according to bottom-up relations, and

(2) it is among the best N (number of visual indices) regions according to bottom-up relations, and

(3) it is better, according to bottom-up relations, than its mutually exclusive regions, and

(4) the presented regions are equal according to top-down attentional relations.

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 (according to bottom-up relations) visual regions exceeds the number of the visual indices;
2) two mutually exclusive visual regions are equally good (according to bottom-up relations);

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

Visual attention is able to resolve the above problems by (1) focusing on only some of the visual regions that are equal according to bottom-up relations, (2) focusing on only one of the mutually exclusive regions, and (3) focusing on region that is worse, according 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 attended to more than others. What is more, the attentional top-down relations can override the influence of bottom-up factors, so a highly attended-to region may be recognized as an object even if it is standing in an unfavorable bottom-up relational structure. By taking these remarks into consideration, the theory of attentional relational construction of visual objects can be stated:

(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 according to top-down attentional relations and those regions are attended to more than all other presented regions, and

(2) *it is attended to more than any of its mutually exclusive regions.*

PREL and *AREL* describe two sets of the 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 standing in a favorable 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 region, even if its bottom-up relations are rather poor.

If a visual region satisfies conditions characterized by PREL or AREL, then it also satisfies the *VO* criterion and is a minimal visual object. Such a region will then gain a visual index; its more complex features (especially its shape) will be processes; it can be grouped with other simple objects to form a more complex one; and finally it will be categorized 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 turn it into an ordinary visual region, which relational characteristics are not recognized.

It should be noted, that according to the above considerations are two distinct expansions of the set of a visual region's structural characteristics and so two different types of representations that are Minimal Visual Objects. However, preattentively constructed objects usually attract attention and also gain favorable top-down relations.

IV. Conclusions

I have argued that the adequate account of visual objects' construction can be provided by a relational notion. According to it, 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, as their presence distinguishes the CSSC shared by tokens of the Minimal Visual Object from the CSSC shared by tokens of the Visual Region, that are maximal non-object representations. The relational notion is coherent with the major figure/ground scientific discrimination models, which claim that regions are distinguished as object by means of comparing their features with those of other regions. More specifically, I have proposed that two types of relational constructing characteristics exist: one connected with the preattentive and a second with the attentive mode of objects' formation.

Apart from introducing the relational notions of visual objects' construction, an additional result show that the most popular notions of objects' structure proposed so far on the ground of analytic metaphysics, connected with unique elements, specific connections, or exemplifying general categories, do not play an important role in describing the constructing characteristics of visual objects. What is more, the analysis suggests that proposals by Clark and Raftopoulos about visual objects' construction need adjustments to accommodate the important role of relational factors.

References

Allaire, E.B. (1963). Bare Particulars. *Philosophical Studies*, 14(1/2), s. 1-8.

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

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.

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, 2000. (IROS 2000)* (vol. 3, pp. 2061-2066). IEEE Conference Publication.

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, 2001. ICCV 2001*, Vol. I, 105-112.

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.

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

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.

Dori, H., & Henik, A. (2003). 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, 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*. Retrieved April 5, 2013, from 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.

Kellman, P. J., & Shipley, T. F. (1991). A Theory of Visual Interpolation in Objects Perception. *Cognitive Psychology*, 23(2), 141-221.

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.

Leung, T., & Malik, J. (1998). Contour Continuity in Region Based Image Segmentation. In H. Burkhardt & B. Neuman (Eds.), *Computer Vision—ECCV' 98, Vol. I, LNCS 1406* (pp. 514-559). Berlin, Heidelberg, Springer Verlag.

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.

Litt, S. (2000). Gestalt Therapy and Humanistic Psychology. *Positive Health*, 53. Retrieved April 5, 2013, from http://www.positivehealth.com/article/holistic-psychotherapy/gestalt-therapy-and-humanistic-psychology.

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

Lowe E. J. (2006). *The Four-Category Metaphysics*, New York: 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, Massachusetts: MIT Press.

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

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

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.

Moore, C. M., Yantis, S., & Vaughan, B. (1998). Object-based visual selection: Evidence From Perceptual Grouping. *Psychological Science*, 9(2), 104-110.

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

Nothdurft, H.-Ch. (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, Massachusetts: 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 K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance* (pp. 1-46). New York, Wiley.

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

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

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

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, Boston: Routledge and Kegan Paul.

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

Tse, P. U. (1999). Complete Mergeability and Amodal Completion. *Acta Psychologica*, 102(2-3), 165-201.

Tsur, R. (1998, September 19). *Metaphor and Figure - Ground Relationship: Comparisons from Poetry, Music, and the Visual Arts* (Figure 2). Retrieved April 5, 2013, from http://www.tau.ac.il/~tsurxx/FigureGround/Figure-ground+mp3New.html.

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.

Vurdlak. (2008, September 22). FAT's "Rubin Vase" Reveals Twin Portraits. In *Mighty Optical Illusions*. Retrieved April 5, 2013, from http://www.moillusions.com/2008/09/fats-rubin-vase-reveals-twin-portaits.html

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.