

Błażej Skrzypulec

Two Types of Visual Objects

Studia Humana nr 14, 26-38

2015

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

Two Types of Visual Objects

Błażej Skrzypulec

Jagiellonian University, Krakow, Poland

e-mail: blazej.skrzypulec@gmail.com

Abstract:

While it is widely accepted that human vision represents objects, it is less clear which of the various philosophical notions of ‘object’ adequately characterizes visual objects. In this paper, I show that within contemporary cognitive psychology visual objects are characterized in two distinct, incompatible ways. On the one hand, models of visual organization describe visual objects in terms of combinations of features, in accordance with the philosophical bundle theories of objects. However, models of visual persistence apply a notion of visual objects that is more similar to that endorsed in philosophical substratum theories. Here I discuss arguments that might show either that only one of the above notions of visual objects is adequate in the context of human vision, or that the category of visual objects is not uniform and contains entities properly characterized by different philosophical conceptions.

Keywords: *perception, metaphysics, visual objects, content, individuation*

Our usual visual phenomenology suggests that the human visual system represents the environment in terms of objects. In various circumstances these “visual objects” can differ in complexity, ranging from shapeless bundles of features that appear at the periphery of the visual field to objects with complex parts and structures that are recognized as exemplars of general categories. Nevertheless, they always seem to be individuals, located in space and possessing certain qualitative properties. The above intuition about the importance of objects in vision is preserved within the majority of scientific models of human perception, which explain, *inter alia*, how figures are discerned from ground [e.g., 26], how objects are represented as being the same despite changes [e.g., 13], or how perceived objects are categorized as exemplars of general types [e.g., 41]. Because of this it seems to be generally accepted that the human vision represents objects.

However, from the philosophical perspective, it seems that that statements claiming that vision represents objects are quite vague. This is because there exist various, mutually inconsistent philosophical notions of objects. In this context it may be asked which conception of objects is correct in the case of visual objects. One of the main demarcation lines separates bundle and substratum theories of objects. According to the bundle theory, objects are identical to relational combinations of features [5], while in substratum theories each object is constituted by an additional element that differs from usual features and is often characterized as an individuator, or as a subject of features [19; 27].

In the paper, I show that in psychological models visual objects are characterized in two distinct ways, one related to the bundle and another related to the substratum conception of objects. More specifically, I argue that the bundle notion of visual objects is connected with models of perceptual organization, while models describing mechanisms that allow us to represent persistence through change characterize visual objects along the lines of substratum theories. Because these two approaches to visual objects are connected with different classes of psychological models, discussion is limited when it comes to whether we should postulate only “bundle” visual objects, only “substratum” visual objects, or if both types are needed to properly characterize the representational abilities of the human visual system.

The main goal of the paper is to sketch a “conceptual map” of the problem, i.e. to investigate the main reasons for stating either that there is only one type of visual object or that we should postulate both “bundle” and “substratum” visual objects. My goal is not to provide a final answer, which in fact cannot be given based on the current state of empirical investigations, but rather to identify conditions whose satisfaction would provide a strong argument for one of the considered options. I start by explicating my usage of the term “visual object”, and then explain the difference between bundle and substratum conceptions of objects (section 1). Subsequently, I show how the distinction between the bundle and substratum conceptions of objects is connected with the treatment of visual objects in models of perceptual organization and models of visual persistence (section 2). The central focus of the paper is discussion of arguments in favour of accepting one or both types of visual objects (sections 3–7).

1. Bundles, substrata, and visual objects

As was stated above, the human visual system seems to represent objects, which may vary in complexity according to their place within the perceptual process. For example, early visual objects may be characterized as combinations of co-located features [40], and later ones as persisting individuals possessing complex parts [33]; while at the highest level of visual information processing, objects can be represented as exemplars of general categories [12]. It is important to note that the characteristics of visual objects do not have to be exactly the same as characteristics of physical objects that causally interact with the visual system. In particular, the same physical object can be represented differently at subsequent stages of the perceptual process. What is more, in cases of misperception, represented characteristics will be different from those possessed by the object that stands in a causal relation with perceptual mechanisms.

Because of this, statements concerning visual objects should be interpreted as concerning the content of visual representation, or, in other words, as specifying the necessary adequacy conditions of visual representations [35]. In the case of human vision, where the category of objects seems to be crucial, adequacy conditions will usually be connected with the presence of certain objects within the visual field. From this perspective, stating that a red square is the visual object of a representation, or, in other words, that a red square is visually represented, means that a necessary adequacy condition of the concerned representation is the presence of a red square in the visual field.

The specification of visual objects as “objects of content”, i.e. objects whose presence is a necessary condition for a visual representation’s adequacy, makes it possible to interpret the claims of the philosophical bundle and substratum theories in a way that allow to treat them as characteristics of representational content. According to the main claims of bundle theories, every object is identical to a relational combination of features [1: 90; 34: 78; 39). In the context of the functioning visual system, this statement can be formulated as a thesis that representing an object is equivalent to representing a proper combination of features (Which combination are the “proper” ones may vary between different bundle theories):

(B) An object is visually represented if and only if a proper combination of features is visually represented.

In other words, thesis (B) states that the necessary adequacy conditions for visual representations representing objects are the same as necessary adequacy conditions for visual representations representing combinations of features. If the thesis (B) is correct, then to characterize the content of visual representations it is sufficient, at least as far as representations of objects are concerned, to describe features and the ways in which they are related, without the need to postulate any additional elements.

Proponents of substratum theories deny the main claim of bundle theories by arguing that objects cannot be identified with combinations of features [9; 18: 140–143). Instead, it is postulated that the structure of every object contains a special element, different from features, often called a “substratum”. This additional element is thought to serve at least one of three roles. First, it may be an individuator, if having the same substrata is an identity criterion of objects [27]. Second, it may be a subject, if it “instantiates” other elements of an object, i.e. features, where “instantiation” is an asymmetric relation [7]. Third, and most common, it may be a unificator, if without the presence of the substratum the other elements of a given structure would not constitute an object [22].

Relying on the main claims of substratum theories—irreducibility of objects to feature-bundles and the presence of a substratum interpreted as an individuator, subject, or unificator—the following thesis can be postulated:

(S) An object is visually represented if and only if a “substratum” element (serving the role of an individuator, subject, or unificator), usually combined with features, is visually represented.

According to (S), and in contrast to (B), the content of visual representations representing objects cannot be sufficiently characterized in terms of combined features. This is because representing an object is always connected with representing a substratum, which serves one of the three roles described above.

In the next section, I show that theses (B) and (S) are assumed within the characteristics of representational content postulated in certain psychological models of vision.

2. Bundles, substrata, and psychological models

One of the most important categories of psychological models of vision are those of perceptual organization [25: 255–257]. Such models describe how the visual system creates a representation of objects in the visual field by representing simpler elements as composing more complex wholes. According to a popular view, connected with classical physiological research by Hubel and Wiesel [11] and Marr’s fundamental work in the cognitive science of vision [21], the human perceptual system starts by representing local discontinuities between different surface features, like brightness or hue. These simple discontinuities, if they stand in proper spatial relations, are combined to form more complex edges, which together compose an early sketch of the visual field.

This sketch, based on the spatial configuration of edges, serves as a starting point for more complex operations. Most importantly, closed edges designate regions filled with uniform surface features, which are regarded as basic units from which visual objects are constructed [26]. First, some “uniform regions” can be distinguished from the ground and thus gain the status of figures, i.e. simple visual objects. This figure-status can be obtained if a region stands in appropriate qualitative relations to its neighbouring regions, where relevant relations are connected with, *inter alia*, being smaller, more convex, or more symmetrical [28; 32]. Second, simple figures, if they are spatially connected, may be combined into complex objects then serving as parts of a higher-level whole [26; 32]. Third, nearby objects possessing similar features can be connected into a perceptual

group, according to some Gestalt-like laws, forming a whole constituted by spatially-dispersed elements [8; 16]. Finally, objects can gain internal structure from the pattern of edges by which they are designated. In particular, points of concavity, at which different borders of an object are close to each other, allow for distinguishing parts within an object's structure [10]. An important subclass of perceptual organization models can be found in conceptions of visual binding, of which the most influential example is the Feature Integration Theory [40]. According to this theory, vision does not start from the representation of local discontinuities, but from representing simple features, without representing any relations between them. Subsequently, due to the work of attentional mechanisms, co-located features forms feature-bundles, which may be regarded as simple visual objects.

Despite the huge variety of perceptual organization models, they all seem to agree on a general view of the nature of visual objects. According to these models, at the beginning of the perceptual process uncombined features, or simple combinations of them (like local discontinuities) are represented, which, by standing in appropriate spatial relations, compose visual objects. Within this approach it seems that representing an object is always equal to representing a relational combination of features. Because of this, models of perceptual organization accept thesis (B), which is a perceptual counterpart of the main postulate of the philosophical bundle theory of objects.

It should be noted that the above conclusion are independent from considerations about the exact point in the perceptual process at which the representational content starts to be present. It may be the case that elements in which the visual system detects local discontinuities are not representations at all, but merely detectors of some causal influences, and the first representations are in fact those of uniform regions or figures distinguished from ground. However, no matter what the earliest elements of content are, representing them, according to models of perceptual organization, will be always equal to representing some combinations of features.

Models of perceptual organization usually focus on stationary phenomena, and lay out the steps of constructing a visual representation of the environment. In this context it is interesting to observe that models whose main focus is to explain how the visual system represents persistence through time assume a different view on visual objects.

Such models of visual persistence postulate perceptual devices, often called "visual indices" or "object files", that are engaged in representing objects but whose role is not to represent common visual features like hue, shape, or localization [13; 17; 31: 37–39). A single index or file is activated when a causal interaction occurs that is, in ordinary circumstances, connected with the presence of an object within the visual field. According to models of visual persistence, indices and files allow us to recognize the numerosity of objects, to track them despite movement and qualitative changes, and to refer to them in order to gain further information regarding their features [30; 36].

The representational content connected with such perceptual devices is limited to representing several numerically distinct individuals and representing whether a presently represented individual is the same as the one represented at an earlier moment. In usual situations, when other perceptual mechanism are also activated, it is not only represented that there are numerically distinct individuals, but also that these individuals possess some features. Because of this, according to models of visual persistence, the structure of visual objects is composed of two types of element: a simple individual and features associated with it.

The above notion of visual objects is inconsistent with thesis (B). Representing an object is not equal to representing a combination of features, but rather to representing a simple individual that may possess some features. Every such individual is numerically distinct from other individuals and so may serve as an individuator of visual objects. It also possesses features and so may be interpreted as a subject. What is more, such individuals seem to be unificators, since other elements of the object's structure, i.e. its features, would not compose an object without being related to an individual. Because of this, models of visual persistence, in contrast to models of perceptual organization, seem to incorporate thesis (S), formulated by relying on the main postulate of substratum theories of objects.

The two incompatible conceptions of visual objects described above—“combinations of features” and “individual plus features”—were developed independently in connection with different classes of psychological models. Because of this, it may be asked whether we really need to postulate both types of visual object in the context of the human vision. In what follows, I investigate different answers to this question and the reasons that underpin them.

3. Simplicity and explanatory power

The bundle notion of visual objects has an initial advantage over the substratum notion because it relies on weaker assumptions. It offers a simpler account of visual content according to which it can be sufficiently described in terms of combinations of features. What is more, it is uncontroversial that features such as colours, shapes, and localizations are represented by the human visual system, and so it seems intuitively plausible that the characterization of visual content consists in specifying arrangements of features. On the contrary, proponents of the substratum approach expand our visual “ontology” by postulating an additional type of element, i.e. substrata, that play some formal roles but cannot be identified with any entities that are usual visually represented. Because of this, the burden of proof seems to be on proponents of the substratum notion of visual objects. If there is no justification for postulating an additional element of visual content in connection with the representations of objects, then the proper account of visual objects will be some version of the bundle approach.

The substratum view of visual objects can be justified by pointing to a perceptual phenomenon, connected with the representational abilities of human vision, that cannot be explained by treating visual objects as relational combinations of features. However, presenting such a phenomenon is no trivial task, since the bundle notion of visual objects, connected to models of perceptual organization, possesses significant explanatory power.

Starting from low-level elements of content, representing local discontinuities may be interpreted as the representation of mutually-exclusive surface features (like different levels of brightness) combined with spatially connected localizations. Further, representing edges may be equal to representing local discontinuities standing in appropriate spatial relations. In similar fashion, a uniform region designated by edges can be characterized as spatially coherent localization that is connected with a surface feature and not a part of a bigger localization connected with the same surface features.

The same approach to characterizing representational content naturally extends to representations connected with further perceptual processes, which allow for figure/ground discrimination and grouping. Representing a figure can be identified with representing a uniform region whose features stand in certain relations to features of neighbouring regions (like “being more convex” or “being more symmetrical”). Processes of grouping are connected with representing figures standing in relations of similarity and spatial proximity; while representing spatially connected figures is often sufficient for representing complex objects composed of them.

It seems that a huge variety of processes responsible for modelling the spatial arrangements of elements in the visual field produce representations whose content can be described in terms of features standing in relations of spatial arrangement, similarity, or comparison (e.g. “being smaller”). In the next section I consider phenomena that may force us to modify this coherent picture and thus make a room for visual objects as characterized by the substratum notion.

4. Synchronic and diachronic individuation

Because one of the main roles of substrata is that of being individuator, it seems plausible that perceptual which consist in representing objects as being the same or different, may justify the application of the substratum notion of visual objects. Amongst phenomena of perceptual

individuation we may distinguish those connected with synchronic individuation, when the visual system differentiates between objects composing a single stationary scene, and those connected with diachronic individuation, when it is recognized whether changing objects are the same or different.

4.1 Synchronic context

In synchronic individuation, the status of features describing the location of objects is controversial. According to some it is impossible to visually represent that two objects simultaneously occupy the same location [4]. However, many authors claim that such situations actually occur in human perception when, for example, overlapping semi-transparent patterns are observed or reflections on glass surfaces are seen [6; 23; 31: 40–42]. If the former are correct and locations play a privileged role in synchronic individuation, then it may be that a visual object x is synchronically identical to a visual object y if and only if x has the same location as y . In the second case, where spatially overlapping objects may be represented as being distinct, a different rule will apply: a visual object x is synchronically identical to a visual object y if and only if x has exactly the same features as y .

Despite this controversy concerning the role of locations, both solutions use only the bundle notion of visual objects. In these cases, representing that there are two objects consists in representing them as having different features concerning localization or representing them as having any other difference in features, and so no new element, like a substratum, is needed. In order to justify applying the substratum notion of visual objects in order to explain synchronic individuation, it would be necessary to show that the visual system can represent completely overlapping objects, which are different while at the same time have share the same localization as well as other features.

I believe that there are two types of perceptual phenomena that can be regarded as those of “complete overlap”. However, in both cases one may raise serious doubts as to whether they really occur in the context of the human vision. First, it can be represented that moving objects overlap at some point in their spatial trajectories. If two such objects are represented as having exactly the same features, such as size, shape, and colour, then during spatial overlap they would be represented as sharing all of their features. If such objects, during overlap, are represented as being distinct, while being represented as having the same features, then they are individuated in virtue of something other than a difference in possessed features. This additional aspect of visual objects serves as a synchronic individuator and can be identified with a substrate. Nevertheless, one may doubt whether such situations of complete overlap during movement are really represented by the human visual system. For example, it may be the case that in every situation of overlap one object is represented as being below the other one due to some depth cues, or that only one object, resulting from merging the two earlier ones, is represented [see 42].

A second type phenomenon that may be regarded as an instance of “complete overlap” is connected with occlusion. It is well-known that an object that hides behind an obstacle and reappears after a short time is visually represented as being the same [38]. What is more, investigations in the field of developmental psychology show that infants can distinguish whether one or two objects have been hidden behind an occluder and represent that these objects still exist during the occlusion [3]. If the visual system is able to represent the presence of two simultaneously occluded objects, then it may seem that representing the difference between them is not founded on representing that they have different features. These objects are represented as having the same approximate localization, connected with the localization of an obstacle, and during the occlusion are not represented as having any features like colour or shape. Again, in this case representing the distinctiveness of objects requires representing something more than an arrangement of features, and this additional “something” can be seen as a substratum. Nevertheless, one may doubt whether situations in which two or more objects are simultaneously represented as being occluded are still

genuine visual representations. Maybe such representations are a product of higher-level reasoning that is not strongly connected with the visual processes.

The above considerations show that the bundle notion of visual objects would be inadequate in fully accounting for perceptual phenomena of synchronic individuation—that is, if there are cases of “complete overlap” in which objects are represented as being distinct while at the same time being represented as having exactly the same features. Such situations may happen when objects overlap during movement or are hidden behind an occluder, but the current state of empirical investigations probably does not allow for deciding whether they occur in the context of human vision. However, if cases of complete overlap really occur, then the representational content they involve should be characterized using the substratum notion of visual objects, in which substrata play the role of individuators.

4.2 Diachronic context

Within the field of cognitive psychology it is widely accepted that the human vision is able to represent objects as being the same despite movement and changes in qualities like colour or shape [see 37]. Because of this, in diachronic individuation it cannot be simply stated, as was plausible in the synchronic cases, that a visual object x is identical to a visual object y if and only if x has exactly the same features as y . An object can be represented as having different features at different times while still being represented as the same individual.

It may seem that representing sameness in the diachronic context automatically leads to the conclusion that the bundle notion of visual objects is inadequate. According to the bundle notion, representing a particular object is equivalent to representing a particular combination of features. If a change occurs, then a different combination of features is represented. So it seems that, according to the bundle notion of visual objects, a different object is also represented. In this case it would be impossible to explain how the visual system represents the diachronic identity of changing objects by treating them as simply relational combinations of features. However, the bundle notion of visual objects does not in fact entail that representing different combinations of features existing at different times cannot be equivalent to representing a single persisting object. For example, it may be claimed that a visual object x is diachronically identical to a visual object y if and only if (I) x stands to y in a similarity-like relation, which is reflexive, symmetric, but not transitive; or (II) x is connected to y by a chain of visual objects standing in such a relation. In this case representing diachronic identity would be equivalent to representing certain relations between combinations of features, so this phenomenon seems to be adequately grasped in terms of the bundle notion of visual objects.

In fact, in psychological works it is frequently claimed that the role of this similarity-like relation is served by spatiotemporal continuity. Results of experiments involving tracking and reidentifying changing objects suggest that continuous movement is usually connected with representing objects as being the same, even if other features, like shape or colour, change [2; 31: 37]. On the other hand, disturbances of spatiotemporal continuity, except cases of occlusion when the object is briefly hidden behind an obstacle, makes object tracking significantly harder and can break the identity of visual objects [38; 20]. Because of this, it may be proposed that representing diachronic identity between objects is equal to representing combinations of features standing in certain patterns of continuity relations.

Nevertheless, the above picture becomes more complicated if we consider ambiguous cases in which the occurrence of spatiotemporal continuity does not determine the pattern of identity relations. Such cases might include splitting-like situations, when a visual object A , at some moment T_1 , is spatiotemporally continuous with two objects B and C at a subsequent moment T_2 . In such a case, if all considered objects have the same features in terms of size, color, etc., three alternative patterns of identity relations are possible. First, object A is not identical to either B or C . Second, object A is identical to both B and C . This would lead to the conclusion that visual

diachronic identity is not the classical identity relation, because characterizing visual identity as transitive would entail a contradiction by identifying objects B and C. The third option is that object A is identical to only one of the objects B and C.

If the third option adequately captures the behaviour of the human visual system in splitting-like cases, then the occurrence of spatiotemporal continuity appears to be insufficient for the identity of visual objects. What is more, such ambiguous situations do not break diachronic identity, because object A is still identified with one of the objects in the subsequent moment. Because of this, representing diachronic identity is not equal to representing continuous bundles of features, and some additional element of content should be postulated that determines whether object A is identified with B or with C. Of course, this additional element may be interpreted as a substratum, serving the role of individuator. However, patterns of visual identity in splitting-like situations have rarely been studied, and there is no decisive account of how the human perceptual system behaves in such situations [see 15; 24].

In summary, the representational content connected with usual cases of synchronic and diachronic visual individuation can be adequately characterized using the bundle notion of visual objects. However, if cases of “complete occlusion” are represented by human vision, or if in splitting-like cases only one “later” object is identified with the “earlier” object, there are serious reasons for applying the substratum notion of visual objects in order to characterize the content connected with these situations.

5. Numerical difference

The perceptual phenomena of individuation in synchronic and diachronic contexts are not the only ones that may justify use of the substratum notion of visual object. In particular, the bundle notion of visual object is clearly inadequate if situations exist in which human vision represents only numerical sameness or difference of objects without representing any of their features. Within such cases, representing an object would be the same as representing a simple individual that is not qualitatively characterized to any extent. Visual objects of this kind can be regarded as the simplest objects satisfying the substratum notion, which are identical with substrata and are not constituted by any features. Of course, such substrata would be individuators as well as unifiers of the visual objects to which they are identical.

According to some models of vision, of which Pylyshyn’s FINST model is probably the most influential example [31: 39–40], experiments involving simultaneous tracking of several objects reveal that objects can be represented as being the same even when we don’t represent any of their features. Similarly, the numerosity of a small set of objects can be immediately grasped without serial counting (a phenomenon known as “subitizing”, [29]), which may suggest that it is possible to represent several numerically different individuals prior to representing their features.

However, it is far from being universally accepted whether the human visual system is actually able to represent such featureless but numerically distinct objects. It is commonly observed that in tracking experiments, changes in targets’ features, such as colour or shape, not only compromise tracking, but participants are also often unable to report seeing these changes [31: 37]. This might show that during tracking objects can be represented as not having any features. Nevertheless, it is harder to argue that objects can be represented without representing their localization. In fact, the proximity of locations is the most important factor determining identity between objects represented at subsequent moments [15].

Because of this, it may seem more plausible that visual representations of objects also always involve representations of their localizations. While accepting this position substantially weakens claims about representing featureless but numerically distinct objects, it might still be sufficient to justify the application of the substratum notion of visual objects. For example, Ronald Rensink, in his works presenting the coherence theory of attention, postulates an early form of visual representation called “layout” [33]. Layout represents some locations as containing objects

without representing any other features of those objects. Layout is thought to serve as a guide for the serial attentional mechanism, which in virtue of layout can quickly gain access to the most important elements of the visual field and then allows us to form more detailed representations of qualitatively rich objects.

A distinguished location represented by layout is a visual object composed of two elements: a feature describing localization and an additional element that differentiates this location from others that do not contain interesting objects. This additional element cannot be identified with any usual visual features, because no such features are represented by layout, and so it can be interpreted as a type of substratum. In this case, the substratum would not serve the role of individuator, due to the fact that different distinguished locations are individualized by the component describing spatial position. However, the substratum within distinguished locations fulfils another usual role, as postulated in the substratum theories. It will be an unificator, because the structure of a distinguished location would not exist without containing a substratum.

The above considerations shows that the perceptual phenomena of individuation are not the only ones that might justify use of the substratum notion of visual object. If the human visual system is able to represent featureless but numerically distinct objects, or to distinguish locations without ascribing any other features to objects, then there will be cases of representational content that cannot be characterized using only the bundle notion of visual objects.

6. No representations without objects

In the two previous sections, I described phenomena whose presence would justify describing visual objects using the substratum notion, but not the bundle notion of visual objects. In addition, it may be asked whether there are reasons to postulate a stronger thesis, namely that the visual objects properly described by the substratum notion are the only ones which are represented by the human visual system. I believe that two such arguments can be found in the literature concerning models of visual persistence.

First, it has been claimed that it is impossible to visually represent combinations of features without representing that they characterize a single object [14: 318–321; 31: 87–89]. To adequately represent the visual field, the perceptual system needs information not only about the presence of certain features but also about their arrangement [4]. However, from the mere fact of representing some features it cannot be inferred which of them coincide. In particular, a colour-feature can, in principle, be combined with many location-features, but combining them randomly would lead to severe misperceptions. To avoid such a result, a colour-feature should not be combined with just any location-feature, but with a feature of proper size, shape, and position within the spatial framework. Since it seems that, from representing a colour-feature, the location-feature with which should it be combined cannot be inferred, one might think that the visual system does not start from the representation of features but from the representation of objects. Further perceptual processes, for example those connected with attentional mechanisms, may allow us to represent the features that this early objects possess. Such objects would be different from usual visual features and thus may be interpreted as substrata. In such a case, there would be no visual object simply identical to combinations of features, since each of them would also be constituted by an additional substratum element.

Second, one may doubt, as mentioned in the section 3, that the early visual processes described in models of perceptual organization play any representational role at all [31: 74–76, 81]. For example, detecting edges or uniform regions may not be connected with representing such patterns, but may consist in merely registering certain causal influences. Only later processes, using this registered information, produce actual representations that possess adequacy conditions and thus also content. If it were the case that the very first representations were connected with mechanism that recognized the numerosity of objects or their diachronic sameness, prior to representing any of their features, then it would be the case that representing objects would always

involve representing numerically distinct individuals that could, but would not have to, possess features.

While the above arguments may have some intuitive appeal, they both assume some rather strong views about the nature of the human vision, which are not sufficiently justified by current empirical results. First of all, even if representing the presence of a certain combination of features has to be preceded by the detection of an object, it is not obvious that such detection has to be accompanied by our representing that there is an additional element, different to these features. It may be the case that processes connected with detecting an object, as well as processes that allow us to decide which of the features should be represented as combined, consist merely in the registration of a certain causal influence and are not connected with the modification of the representational content.

According to the second argument, it should be observed that a crucial feature of representations is the possibility of misrepresentation. If something cannot be inadequate, such as the effect of a causal influence that is not “right” or “wrong”, but is just determined by the properties of the elements engaged in an interaction, it is not a representation. While it may be the case that simple operations described in models of perceptual organization, like those responsible for detecting edges, are not representational but just register causal influences, this is less likely in the case of grouping and distinguishing figures from a ground. To represent some elements as forming a single perceptual group it has to be judged that they are significantly similar. Analogously, to represent a certain region as a figure distinguished from a ground its features have to seem more important than those of neighbouring regions. The effects of both these operations may intuitively seem prone to error and not automatically determined by the properties of stimuli. In addition, the second argument assumes that human vision may represent numerically different objects without representing their features, which is also far from being obvious (see section (5)).

7. Bundles, substrata, or both?

Based on the above investigations, I shall now try to characterize the conditions in which it would be justified to postulate the existence of one or both types of visual object in the context of human vision. More specifically, three hypotheses may be considered:

(I) All visual objects satisfy the bundle notion, and so, according to thesis (B), representing an object is equivalent to representing a relational combination of features.

(II) All visual objects satisfy the substratum notion, and so, according to thesis (S), representing an object is equivalent to representing a substratum that may be combined with some features.

(III) Some visual objects satisfy the bundle notion and some satisfy the substratum notion, i.e., in the context of human vision there are two types of visual object with different ontological structures.

The considerations presented in sections 3 to 6 reveal five questions that are relevant in judging hypotheses (I), (II), and (III):

(1) Are there cases in which human vision represents distinct objects that share all their features (that is, situations of “complete overlap”)?

(2) Is it the case that in splitting-like situations diachronic identity is maintained between the earlier object and one of the later objects?

(3) Are there cases in which human vision represents numerically distinct objects without representing any of their features?

(4) Are there cases in which human vision represents locations as containing objects without representing any other features of these objects?

(5) Is it the case that a combination of features cannot be represented without representing an element that is different from usual features and serves one of the usual roles of substrata?

To justify hypothesis (I), we have to answer ‘no’ to all the above questions. The phenomena referred to in questions (1)–(4) involve visual objects containing elements that are different to features; as such, these visual objects cannot be adequately characterized using the bundle notion. What is more, a positive answer to question (5) would entail that there are no visual objects that are simply combinations of features.

The last question is also crucial for hypothesis (II). Only if we answer ‘yes’ do all visual objects satisfy the substratum rather than the bundle notion. The other questions are not relevant for hypothesis (II), since answering them positively leads to the applicability of the substratum notion in some, but not all cases in which objects are visually represented. If hypothesis (II) is true, then all visual objects are constituted by a substratum playing the role of unificator; but this substratum is not necessarily an individuator, since objects may be individuated in virtue of possessing different arrangements of features.

Finally, hypothesis (III) would be justified if the answer to at least one of questions (1)–(4) was positive, but the answer to question (5) negative. In such a case, there would be some phenomena involving visual objects that could not be reduced to combinations of features, but there would be no reason to suppose that there were no visual objects satisfying the bundle notion. In the case of phenomena related to questions (1)–(3), the substrata constituting visual objects would be both unificators and individutors. However, in the case of distinguished locations referred to in question (4), substrata would only serve as unificators, because such visual objects are individuated by features describing locations. It is worth noting that none of the discussed phenomena, whose occurrence would justify the presence of visual objects satisfying the substratum notion, suggest a need to postulate substrata serving the role of subjects of features. This is because there is no salient reason to assume that the relation between substrata and features in the structure of visual objects has to be asymmetric.

In earlier sections, I claimed that the current state of research concerning vision prevents us from determining which of hypothesis (I)–(III) is true for human perception. Nevertheless, the investigations conducted within this paper suggest that hypothesis (II) may be least probable, since it relies on strong and controversial assumptions about the nature of visual processing; and, by contrast, the bundle notion of visual objects possesses significant explanatory power. However, to reject hypothesis (I) it would be sufficient to prove the existence of a single phenomenon, related to individuation or the representation of numerical difference, like those described in sections 4 and 5, such that the representational content connected with it could not be fully characterized in terms of a combination of features. Because of this, one may rationally suppose that the category of visual objects is not ontologically uniform, but contains two types of objects adequately described by the bundle or substratum theories respectively.

8. Conclusion

I argued that scientific models of human vision assume two, mutually inconsistent notions of visual objects. Models of perceptual organization adopt the main thesis of the philosophical bundle theory of objects and so characterize visual objects as relational combinations of features. Models of visual persistence, however, characterize visual objects in accordance with the philosophical substratum theory of objects and because of this characterize visual objects as constituted by a substratum that cannot be identified with any usual features.

After investigating reasons for claiming the existence of only one or both types of visual object in the context of human vision, I argued that it would be difficult to defend the thesis that all visual objects satisfy the substratum notion. However, it is not necessarily the case that every visual object can be adequately described as a combination of features. If within the context of human vision certain phenomena occur, connected, *inter alia*, with the individuation of objects or with representing distinct objects in the absence of features, then the substratum notion is required for

their characterization. In this case, the category of visual object would contain two types of ontologically different structures.

References

1. Armstrong, D. M. 1978, *Universals and Scientific Realism. Vol. I: Nominalism and Realism*. Cambridge: Cambridge University Press.
2. Bahrami, B. 2003, "Object Property Encoding and Change Blindness in Multiple Object Tracking", *Visual Cognition* 10, pp. 949–963.
3. Carey, S., Xu, F. 2001, "Infants' Knowledge of Objects: Beyond Object Files and Object Tracking", *Cognition* 80(1–2), pp. 179–213.
4. Clark, A. 2004, "Sensing, Objects, and Awareness: Reply to Commentators", *Philosophical Psychology* 17(4), pp. 563–589.
5. van Cleve J. (1985), *Three Versions of the Bundle Theory*, *Philosophical Studies*, 47(1), 95–107.
6. Cohen, J. 2004, "Objects, Places, and Perception", *Philosophical Psychology* 17(4), pp. 471–495.
7. Denkel, A. 2000, "The Refutation of Substrata", *Philosophy and Phenomenological Research* 61(2), pp. 431–439.
8. Elder, J. H., Goldberg, R. M. 2002, "Ecological Statistics of Gestalt Laws for the Perceptual Organization of Contours", *Journal of Vision* 2(4), pp. 324–353.
9. Hochberg, H. 1965, "Universals, Particulars, and Predication", *The Review of Metaphysics* 19(1), pp. 87–102.
10. Hoffman, D. D., Richards, W. A. 1984, "Parts of Recognition", *Cognition* 18(1–3), pp. 65–96.
11. Hubel, D., Wiesel, T. N. 1962, "Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex", *The Journal of Physiology* 160, pp. 106–154.
12. Hummel, J. E. (2013), *Object Recognition*, in D. Reisburg (ed.), *Oxford Handbook of Cognitive Psychology*, Oxford University Press, Oxford, pp. 32–46.
13. Kahneman, D., Treisman, A. M., Gibbs, B. J. 1992, "The Reviewing of Object Files: Object-Specific Integration of Information", *Cognitive Psychology* 24(2), pp. 175–219.
14. Keane, B. P. 2009, *Visual Objects as the Referents of Early Vision: A Response to A Theory Of Sentience*, in D. Dedrick, L. Trick (eds.), *Computation, Cognition, and Pylyshyn*, The MIT Press, Cambridge, MA, pp. 303–333.
15. Keane, B. P., Pylyshy, Z. W. 2006, "Is Motion Extrapolation Employed in Multiple Object Tracking? Tracking as a Low-Level, Non-Predictive Function", *Cognitive Psychology* 52(4), pp. 346–368.
16. Kubovy, M., Holcombe, A. O., Wagemans, J. 1998, "On the Lawfulness of Grouping by Proximity", *Cognitive Psychology* 35(1), pp. 71–98.
17. Leslie, A. M., Xu, F., Tremoulet, P. D., Scholl, B. J. 1998, "Indexing and the Object Concept: Developing 'What' and 'Where' Systems", *Trends in Cognitive Science* 2(1), pp. 10–18.
18. Loux, M. J. 1978, *Substance and Attribute. A Study in Ontology*. Dordrecht: D. Reidel.
19. Lowe, J. E. 2000, "Lock, Martin, and Substance", *The Philosophical Quarterly* 50(201), pp. 499–514.
20. van Marle, K., Scholl, B. J. 2003, "Attentive Tracking of Objects vs. Substances", *Psychological Science* 14, pp. 498–504.
21. Marr, D. 2010, *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*. Cambridge, MA: MIT Press.
22. Martin C. B. (1980), "Substance Substantiated", *Australasian Journal of Philosophy* 58(1), pp. 3–10.
23. Matthen, M. P. 2004, "Features, Places, and Things: Reflections on Austen Clarks's Theory of Sentience", *Philosophical Psychology* 17(4), pp. 497–518.
24. Mitroff, S.R., Scholl, B.J. Wynn, K. 2004, "Divide and Conquer: How Object Files Adapt When a Persisting Object Splits Into Two", *Psychological Science* 15, pp. 420–425.

25. Palmer, S. E. 1999, *Vision Science: Photons to Phenomenology*. Cambridge, MA: MIT Press.
26. Palmer, S., Rock, I. 1994, "Rethinking Perceptual Organization: The Role of Uniform Connectedness", *Psychonomic Bulletin and Review* 1(1), pp. 29–55.
27. Park, W. 1990, "Haecceitas and the Bare Particular", *The Review of Metaphysics* 44(2), pp. 375–397.
28. Pomerantz, J. R., Kubovy, M. 1986, *Theoretical Approaches to Perceptual Organization. Simplicity and Likelihood Principles*, in K. R. Boff, L. Kaufman, J. P. Thomas (eds.), *Handbook of Perception and Human Performance*, Wiley, New York, pp. 1–46.
29. Pylyshyn, Z. W. 1994, "Some Primitive Mechanisms of Spatial Attention", *Cognition* 50, pp. 363–384.
30. Pylyshyn, Z. W. 2001, "Visual Indexes, Preconceptual Objects, and Situated Vision", *Cognition* 80(1), pp. 127–158.
31. Pylyshyn, Z. W. 2007, *Things and Places. How the Mind Connects with the World*. Cambridge, MA: MIT Press.
32. 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), pp. 155–166.
33. Rensink, R. A. 2000, "The Dynamic Representation of Scenes", *Visual Cognition* 7(1/2/3), pp. 17–42.
34. Russell, B. 2009, *Human Knowledge: Its Scope and Limits*. New York: Routledge.
35. Schellenberg, S. 2011, "Perceptual Content Defended", *Nous* 45(4), pp. 714–750.
36. Scholl, B. J. 2001, "Objects and Attention: The State of Art", *Cognition* 80(1), pp. 1–46.
37. Scholl, B. J. 2007, "Object Persistence in Philosophy and Psychology", *Mind and Language* 22(5), pp. 563–591.
38. Scholl, B. J., Pylyshyn, Z. W. 1999, "Tracking Multiple Items through Occlusion: Clues to Visual Objecthood", *Cognitive Psychology* 38, pp. 259–290.
39. Simons, P. 1994, "Particulars in Particular Clothing: Three Trope Theories of Substance", *Philosophy and Phenomenological Research* 54(3), pp. 553–575.
40. Treisman, A. M. 1998, "Feature Binding, Attention and Object Perception", *Philosophical Transactions of the Royal Society of London. Series B. Biological Sciences* 353(1373), pp. 1295–1306.
41. Ullman, S. 1996, *High-Level Vision. Object Recognition and High-Level Vision*. Cambridge, MA: MIT Press.
42. Viswanathan, L., Mingolla, E. 2002, "Dynamics of Attention in Depth: Evidence From Multi-Element Tracking", *Perception* 31(12), pp. 1415–1437.