# What is a visual object?

**Jacob Feldman**

Department of Psychology, Center for Cognitive Science, Rutgers University, New Brunswick, NJ 08903, USA

**The concept of an 'object' plays a central role in cognitive science, particularly in vision, reasoning and conceptual development – but it has rarely been given a concrete formal definition. Here I argue that visual objects cannot be defined according to simple physical properties but can instead be understood in terms of the hierarchical organization of visual scene interpretations. Within the tree describing such a hierarchical description, certain nodes make natural candidates as the 'joints' between objects, representing division points between parts of the image that cohere internally but do not perceptually group with one another. Thus each subtree hanging from such a node corresponds to a single perceived 'object'. This formal definition accords with several intuitions about the way objects behave.**

Objects are everywhere in cognitive science. Objects are thought to be the building blocks of children's conception of the physical world [1,2]; to delineate the boundaries respected by visual attention [3–5]; and to influence neural processing even at the earliest stages of visual cortex [6,7].

But what exactly *is* an object? In a phrase due to the American jurist Potter Stewart, 'we know one when we see one' – but what does the word actually mean? How do we know where one object ends and the next begins?

The premise of this article is that this is not (as it were) an 'objective' question, but rather one that relates to how we mentally divide the world up into coherent units. That is, it is less about physics and more about the mental assumptions lurking behind the word 'coherent'. The division of the world into objects seems so intuitive and effortless, at least under everyday conditions, that we speak about this division as if the world provided it overtly, without any contribution from our brains. But if cognitive science has shown anything, it has shown that what seems subjectively obvious is often the result of complex and subtle computations. The division of the world into objects is a case in point. The fallacy (philosophers would call it 'Naive Realism') is epitomized by Woody Allen's tale of the Great Roe, a mythical beast with 'the head of a lion, and the body of a lion, although not the same lion' [8]. If (and only if) it looks like an object, and quacks like an object, it's an object. Perception dictates.

In the same way that a fist is something that a set of five fingers turns into only when they are organized a particular way, objects are subsets of the world to which has been attached – by the perceiver – a particular kind of subjective organization. And like a fist, objects take on special significance and definite properties only by virtue of this organization. Therefore, in seeking a definition of objects, I would argue, we need to focus not on how the *world* is structured, but rather on how our subjective perceptual *interpretations* are organized, and then ask how this kind of organization most naturally decomposes into object-like components.

But what kind of organization turns inchoate visual 'stuff' into a coherent object? Unfortunately, no simple answer to this question is to be found in the literature on perceptual grouping.

First, no *single* grouping cue defines objects. 'Real' objects and object boundaries tend to obey a variety of nice properties – including closure [9,10], connectedness [11], convexity [12,13], good continuation in their contours [14–18], regularity of shape [19,20], and so forth. But although each of these properties contributes to the perception of objects, none is, in and of itself, essential. Counterexamples can be found for each principle – perfectly good objects that are concave, have irregular shapes or boundaries, and so on. Similarly, one can easily find non-objects obeying any one of the principles – for example, crossed sticks (which are uniformly connected and yet perceived as multiple objects).

Hence some more abstract organizational principle is required, perhaps drawing together several otherwise disparate individual rules. The Gestalt term *Prägnanz* – 'goodness of form' – is one famous attempt at such a principle, as are other terms arising from the developmental literature, such as 'boundedness' and 'cohesiveness' (e.g. see [1]). These terms are intuitively helpful but, I think, too vague or ill-defined to get us beyond Stewart's 'we know one when we see one'.

Second, although visual psychologists often loosely use the term 'objects' to refer to the fruits of grouping processes, most grouping processes studied in the literature relate to the formation of contours, textures and surfaces – image units that cohere but are not, by themselves, complete objects. Objects are in a sense the ultimate product of such processes, carried out until there is no more grouping left to be done. As Anne Treisman put it [21], objects are 'complex wholes'. But what does it mean for a visual group to be 'whole?'

## Hierarchical organization in vision

Many researchers have noted that perceptual organization tends to be hierarchical, with spatial relations defined at different spatial scales. Such proposals have come out of psychology [10,22–24], as well as computer vision [25]. Formally, a hierarchical description corresponds

*Corresponding author:* Jacob Feldman (jacob@ruccs.rutgers.edu).

http://tics.trends.com 1364-6613/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved. doi:10.1016/S1364-6613(03)00111-6

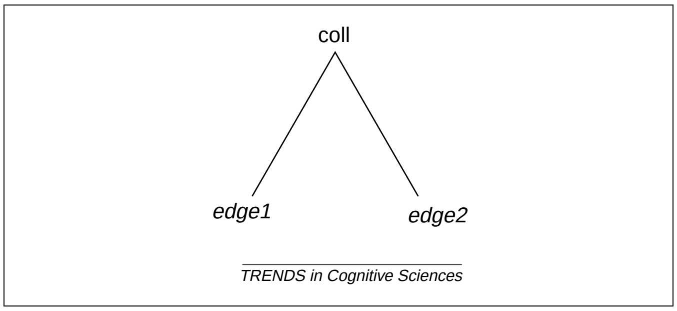


**Fig. 1.** A simple tree (see text).

to a 'tree', in which the root node (normally drawn at the top) describes the configuration at the most global level, while its subtrees describe finer or more local spatial relations, and their subtrees even finer ones, all the way down to the leaves, which correspond to individual visual elements (pixels, dots, oriented edges, etc.). Recently, this type of organization has become popular in computer vision (e.g. see [26,27]). In these proposals, visual groups correspond to particular components (subtrees) of the tree, namely those that are in a formal sense maximally disjoint from each other. Grouping then becomes a process of finding the cleanest 'joints' in the tree, points technically referred to as 'normalized cuts'. The proposal below is in very much the same spirit, although expressed in somewhat more general terms.

In general then, at each node in such a tree is some sort of representation of the spatial relations in force among the tree's subtrees. For example, the tree in Fig. 1 describes a collinear arrangement of two visual elements *edge1* and *edge2*. The representation at each node is expressed in your visual representation language of choice, probably using the regularities and principles mentioned above (closure, collinearity, connectedness, convexity, etc.) – depending on your chosen theory of visual representation. As we are focusing here on the more abstract principle drawing these individual rules together, for current purposes the choice doesn't much matter. (For the purpose of producing a useful description of the scene, the choice matters a lot of course.) The idea is simply that visual items that are the common arguments of a visual predicate – that is, both children of the same node in the tree – tend to be grouped [28].

We also need some way of representing the 'degree of regularity' of the spatial arrangement between the subtrees, with a special way of designating zero regularity. Zero regularity means 'no special relationship' – what mathematicians call 'general position' or a 'generic' relationship. Three points are said to be in general position if they do not fall in a line (that would be a 'special' configuration), four points are in general position if they do not fall in a plane, and so forth. 'Special configurations' (what generic configurations are *not*) are defined by the chosen language of spatial representations, and include visual relationships that the system chooses to elevate to the status of 'atomic' descriptors – such as the grouping predicates listed above – as well as any possible combination of these atoms. So regular (non-generic, special) means 'exhibiting some structure recognized by the system, such as collinear, parallel, touching… [or whatever]', and generic means 'none of the above'.

In other articles [29,30] I have proposed ways of measuring the degree of regularity numerically, with generic corresponding to zero. (The simplest way is simply to count the descriptors that apply, which gives the 'logical depth' of the configuration.) The details are not important here: the point is that visual elements bearing some special, perhaps non-accidental, relation to each other tend to be grouped.

How does the system actually compute the best hierarchical description of an image? That is a separate and much larger question, which is taken up at length elsewhere [29,30]; see also [26,27]). Here, I focus instead on the question of how, given a tree, one can most reasonably divide it into coherent and self-contained units – or objects.

## Disjoints

Within the entire tree that hierarchically represents a visual scene, a special role is played by nodes that are generic (zero structure), denoted with the symbol Ø (see Box 1 and Fig. 2). At such nodes, the subtrees have literally

### Box 1. Key terms

A **tree** is a diagram with a hierarchically branching structure, usually drawn with the 'root' at the top (see Fig. I). The nodes adjoining a node below are its **children**. The node at the top of the tree is called the **head**, and those at the very bottom are called the **leaves**.

A **subtree** is a tree contained within a tree, including all the nodes and edges that branch downwards from any given node (the head of the subtree).

Trees are useful for depicting the **spatial relations** among elements and groups of elements of a visual image. The leaves of the tree correspond to individual visual elements, such as dots or edges, and subtrees correspond to perceptual groups at various levels of the hierarchy.

Spatial relations depicted at each node can include **non-accidental properties** (i.e. spatial relations that are unlikely to be an "accident", such as collinearity, parallelism, etc). When there is no such special spatial relationship, the relation is called **generic** ('nothing special'); otherwise, it is **non-generic** or **regular**. Generic nodes are indicated by a Ø in the tree diagram.

A **disjoint** is a generic node with at least one regular child.

A **disjoint regular subtree**, or an **'object'**, is a subtree that (a) hangs from a disjoint and (b) has a regular head.

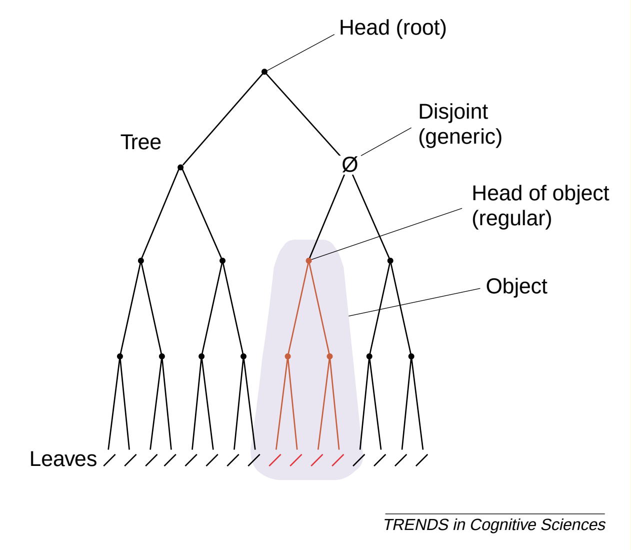


**Fig. I.** Illustration of terms (see text).

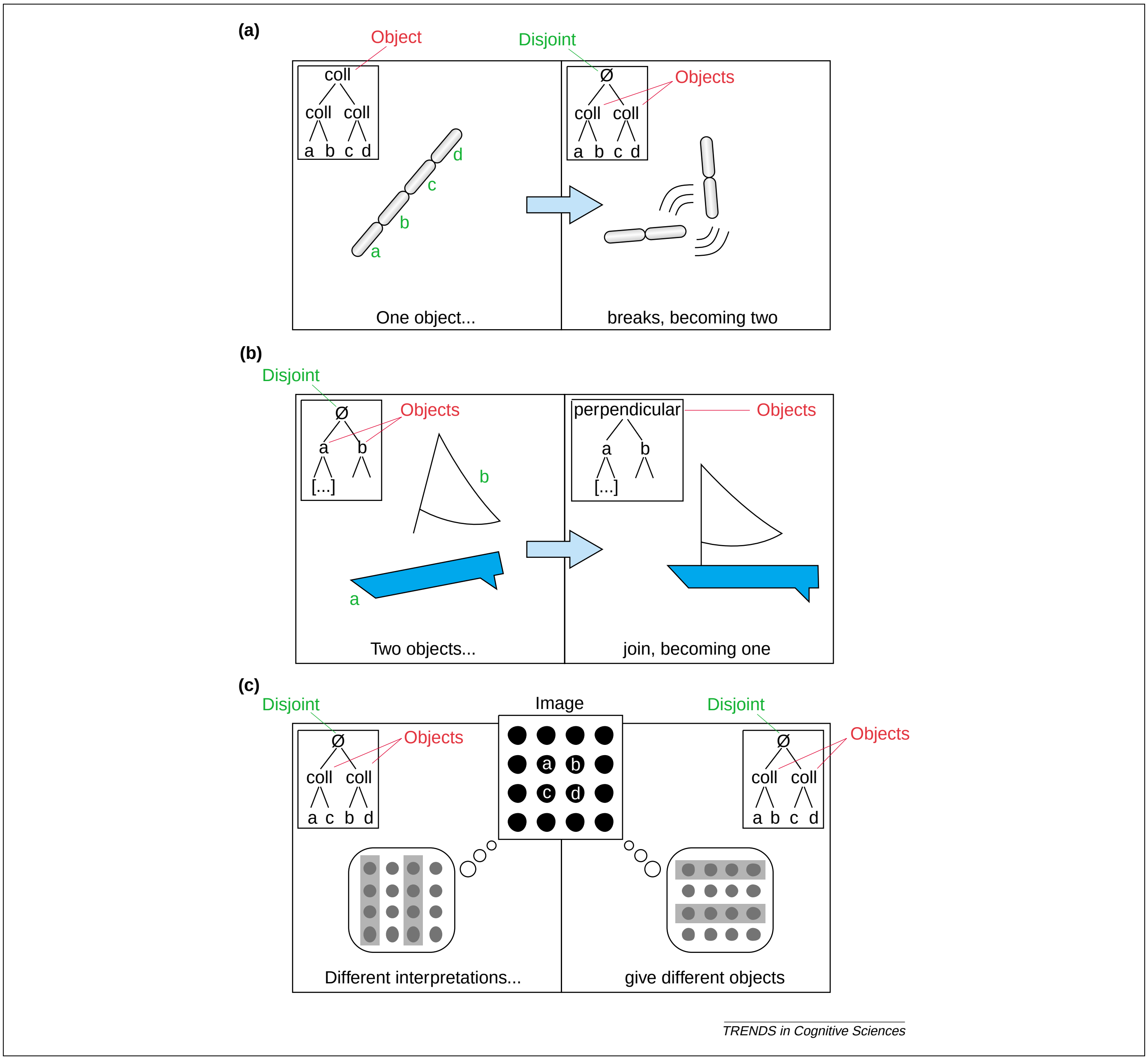


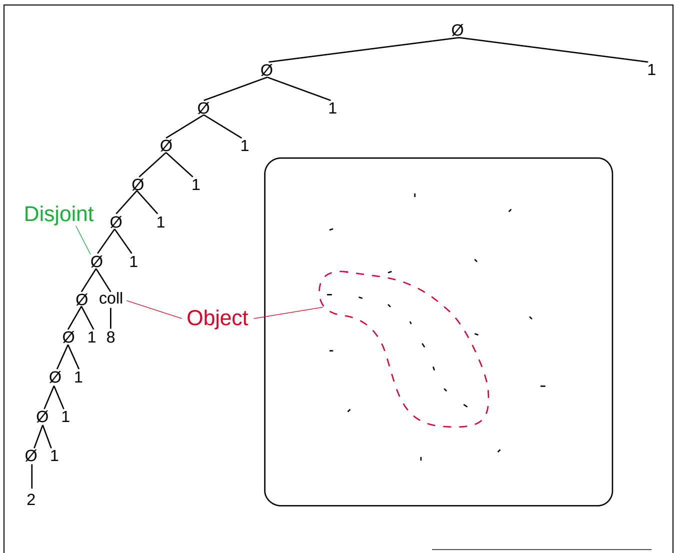
**Fig. 2**. Three consequences of the object definition. (a) When an object breaks, with an accidental spatial relation replacing a non-accidental one, the corresponding interpretation tree forms a disjoint and splits into two objects. (b) When two objects are juxtaposed non-accidentally, a disjoint disappears and two objects coalesce into one. (c) When an image (center) can be interpreted two different ways (left and right), phenomenal objects can change. Here, the grid of dots can be seen as horizontally or vertically striped.

no recognizable mutual spatial relation, and hence are not grouped together at all. Sometimes, these subtrees can themselves be structureless – for example only having generically related visual items in their scope (e.g. random dots).But when they are not – when their top node is *non-generic* – then they constitute complete formal substructures of the visual hierarchy. Such subtrees are both ‘disjoint’ (i.e. they hang from a disjoint node, and thus are not grouped with the rest of the image) and ‘regular’ (i.e. their head node is non-generic) – so I will refer to them as ‘disjoint regular subtrees’, and call the generic nodes from which they hang ‘disjoints’ (see Box 1). The disjoint regular subtrees, separated by disjoints, are the largest internally coherent components that do not group with the rest of the tree. Hence in a very natural sense, the disjoint regular subtrees are the objects [31].

Note that it is a consequence of the formal structure of trees that each subtree is connected to the rest of the tree at only one place – its top node. (By definition, trees have diagrams that only branch, never rejoin, as you move down.) Hence ‘breaking’ this node is enough to completely disconnect the subtree from the rest of the tree. The disjoint regular subtrees – objects – fall off the tree like ripe plums clipped at their stems.

## Tweaking the definition

We can easily weaken the above object definition to allow for less-than-perfect disjunction of objects from the rest of



TRENDS in Cognitive Sciences

**Fig. 3**. A simple configuration of 20 edges, seeming to contain an object and some random background texture, and a tree description of it (see [29] for an explanation of how the tree was computed). The single disjoint and object are indicated. In the tree, Ø denotes a generic node, and the numbers indicate how many image elements are in the scope of each node. Hence the coll/8 subtree indicates a collinear arrangement of 8 edges, which is the object visible in the middle of the image.

the scene. If our regularity-measuring function gives numeric ratings of the regularity of nodes, then a useful measure of 'degree of objecthood' is simply the numeric difference in regularity between a subtree and its parent node. This number will be larger the more internally coherent the subtree is compared with how strongly it is bound to other subtrees. Thus, the handle of a mug might not count as a perfect 'object' by the pure definition because it does have a non-accidental binding to the rest of the mug – their relation is not a true disjoint – but it scores high on the degree-of-objecthood measure because its spatial relation to the mug is weaker than its internal spatial relations.

Note that not every node in a tree is part of *any* disjoint regular subtree – that is to say that not every piece of visual stuff congregates into an object. If an image consists of a coherent structure surrounded by a mass of random texture, the corresponding formal interpretation will be a disjoint regular subtree plus a lot of other generic nodes (with no regular subtrees hanging from them) – which means one object plus a lot of stuff that isn't part of any object. Of course, this corresponds exactly to the intuitive interpretation of 'figure plus background.' Figure 3 shows a computational example.

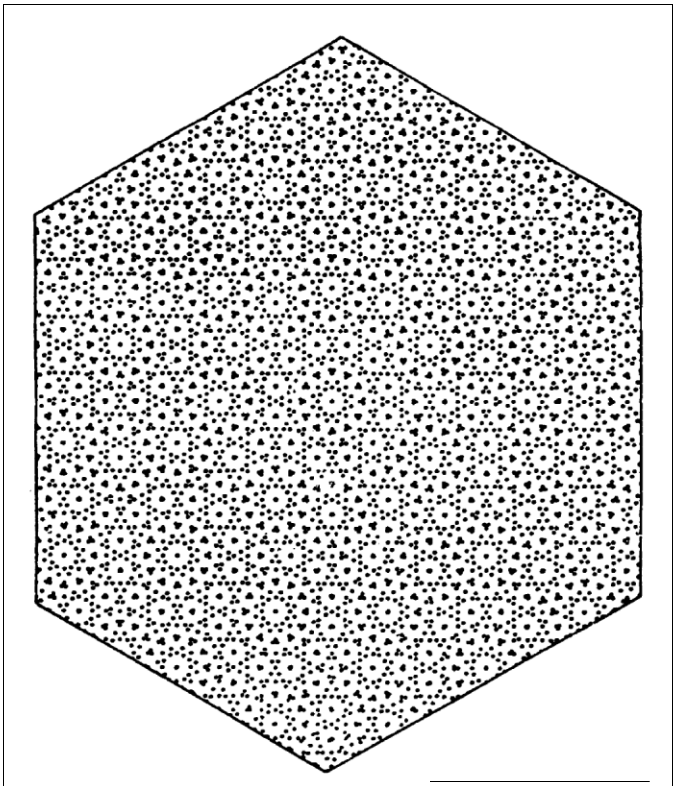
## Some consequences of the object definition

This definition of objects has several obvious consequences, all of which neatly match everyday intuitions about objects. Figure 2 gives illustrations, including explicit schematic diagrams of the corresponding trees, showing how disjoints form and dissolve when spatial relations change.

(a) When a coherent object breaks in two, with the two parts taking on a generic relationship – for example, randomly skewed – then the new configuration will be cognized as two objects. This new generic relation becomes a disjoint, replacing the node that previously sat at the top of the entire tree (Fig. 2a).

(b) When two distinct objects are affixed to each other in a non-accidental way, they become cognized as one object (Fig. 2b).

(c) Different ways of subjectively organizing the scene can lead to different 'objects' (Fig. 2c). This phenomenon, beautifully illustrated by a multistable figure due to Marroquin (Fig. 4), epitomizes the subjective nature of object organization. The objects in Marroquin's figure are not 'real'; they are ephemeral elements of our interpretation – I would argue, disjoint regular subtrees fluidly changing as the overall tree is continually reorganized. The same is equally true for everyday objects like chairs and pencils, no matter how stable *their* interpretation trees are by contrast.



TRENDS in Cognitive Sciences

**Fig. 4**. A figure due to Jose L. Marroquin (reproduced from [35]). How many objects are there?

An additional, more subtle, prediction stems from the degree-of-objecthood measure. It is well known that objects exhibit several attentional effects, for example a response-time benefit that accrues to certain comparisons within their borders [32,33]. If the reasoning given above is correct, such object benefits should *increase* monotonically with the difference in regularity between a subtree and its parent node. Recent experiments in my laboratory [34] have confirmed this prediction in detail: the object benefit increases steadily with the degree of regularity of a line configuration. Regularity induces objecthood.

## Conclusion

Objects cannot be adequately defined by any simple physical property, nor even any simple *perceptual* property; they require a more abstract definition. Intuitively, objects are components of the subjective visual interpretation that are both coherent and complete. Hence defining them formally means asking how subjective interpretations are formally structured, and then considering how this kind of formal structure most naturally decomposes, or 'breaks apart at the seams'. A very natural way of expressing perceptual interpretations is in terms of hierarchical descriptions, or trees. Then the only question is how to divide the whole tree most naturally into subtrees, and I have argued that the most natural choice is into disjoint regular subtrees. Each such subtree is a component of the tree that contains internal structure, but is formally disjoint from the rest of the tree. This very naturally captures the intuition of a visual unit that coheres, but is not grouped with other units in the image – an object.

## Future directions

The question of objects is, to be sure, broader than that discussed here. Objects are important not only because of their role in perceived spatial organization, which I have focused on, but also because of their role in our conceptual organization of the world. Consider these three different roles played by objects, each of which has been studied in the literature:

(1) Objects are the units of our perceived physical world – spatially coherent bundles of visual stuff (the sense of 'object' explored in this article);
(2) Objects are the units of our ontology – the things we think of as having independent existence, properties, and attributes;
(3) Objects are the units of mental dynamics – the things we think of as having fixed existence in a world that changes over time.

These three senses of 'object' are substantially logically independent from each other. For example it is not logically necessary to assume that only spatially coherent units tend to be stable over time, or that properties should be associated only with spatially coherent units. Spatial coherence, material properties, and dynamics are all independent – in principle, but *not* in our conception of things. Hence in my view the question that ought to drive future research on objects is: what do (1), (2), and (3) have to do with each other?

**Acknowledgements**

I am grateful to Randy Gallistel, Zenon Pylyshyn, Manish Singh, Johan Wagemans, and the anonymous reviewers for helpful comments and discussions. Preparation of this manuscript was supported by NSF SBR-9875175.

## References

1 Spelke, E.S. (1990) Principles of object perception. *Cogn. Sci.* 14, 29–56
2 Xu, F. and Carey, S. (1996) Infants' metaphysics: the case of numerical identity. *Cogn. Psychol.* 30, 111–153
3 Baylis, G.C. (1994) Visual attention and objects: two-object cost with equal convexity. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 208–212
4 Duncan, J. (1984) Selective attention and the organization of visual information. *J. Exp. Psychol. Gen.* 113, 501–517
5 Scholl, B. *et al.* (2001) What is a visual object? Evidence from target merging in multiple object tracking. *Cognition* 80, 159–177
6 Albright, T.D. (1994) Why do things look as they do? *Trends Neurosci.* 17, 175–177
7 Olson, C.R. (2001) Object-based vision and attention in primates. *Curr. Opin. Neurobiol.* 11, 171–179
8 Allen, W. (1972) *Without Feathers*, Ballantine Books
9 Elder, J. and Zucker, S. (1993) The effect of contour closure on the rapid discrimination of two-dimensional shapes. *Vision Res.* 33, 981–991
10 Pomerantz, J.R. *et al.* (1977) Perception of wholes and their component parts: some configural superiority effects. *J. Exp. Psychol. Hum. Percept. Perform.* 3, 422–435
11 Palmer, P. and Rock, I. (1994) Rethinking perceptual organization: the role of uniform connectedness. *Psychon. Bull. Rev.* 1, 29–55
12 Bertamini, M. (2001) The importance of being convex: an advantage for convexity when judging position. *Perception* 30, 1295–1310
13 Liu, Z. *et al.* (1999) The role of convexity in perceptual completion: beyond good continuation. *Vision Res.* 39, 4244–4257
14 Caelli, T.M. and Umansky, J. (1976) Interpolation in the visual system. *Vision Res.* 16, 1055–1060
15 Feldman, J. (1997) Curvilinearity, covariance, and regularity in perceptual groups. *Vision Res.* 37, 2835–2848
16 Foster, D.H. (1979) Discrete internal pattern representations and visual detection of small changes in pattern shape. *Percept. Psychophys.* 26, 459–468
17 Pizlo, Z. *et al.* (1997) Curve detection in a noisy image. *Vision Res.* 37, 1217–1241
18 Smits, J.T. and Vos, P.G. (1987) The perception of continuous curves in dot stimuli. *Perception* 16, 121–131
19 Feldman, J. (2000) Bias toward regular form in mental shape spaces. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1–14
20 Kanizsa, G. (1979) *Organization in Vision: Essays on Gestalt Perception*, Praeger Publishers
21 Treisman, A. (1986) Properties, parts and objects. In *Handbook of Perception and Human Performance: Cognitive Processes and Performance* (Vol. 2) (Boff, K.R., ed.), pp. 35-1–35-70, John Wiley and Sons
22 Baylis, G. and Driver, J. (1993) Visual attention and objects: evidence for hierarchical coding of location. *J. Exp. Psychol. Hum. Percept. Perform.* 19, 451–470
23 Geisler, W.S. and Super, B.J. (2000) Perceptual organization of two-dimensional patterns. *Psychol. Rev.* 107, 677–708
24 Palmer, S.E. (1977) Hierarchical structure in perceptual representation. *Cogn. Psychol.* 9, 441–474
25 Marr, D. and Nishihara, H.K. (1978) Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. Lond. Ser. B* 200, 269–294
26 Amir, A. and Lindenbaum, M. (1998) A generic grouping algorithm and its quantitative analysis. *IEEE Trans. Patt. Anal. Mach. Intell.* 20, 168–185
27 Shi, J. and Malik, M. (2000) Normalized cuts and image segmentation. *IEEE Trans. Patt. Anal. Mach. Intell.* 22, 888–905
28 Pomerantz, J.R. and Pristach, E.A. (1989) Emergent features, attention, and perceptual glue in visual form perception. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 635–649
29 Feldman, J. (1997) Regularity-based perceptual grouping. *Comput. Intell.* 13, 582–623
30 Feldman, J., Perceptual grouping by selection of a logically minimal model. *Intl. J. Comput. Vis.* (in press)
31 Feldman, J. (1999) The role of objects in perceptual grouping. *Acta Psychol. (Amst.)* 102, 137–163
32 Behrmann, M. *et al.* (1998) Object-based attention and occlusion: evidence from normal participants and a computational model. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 1011–1036
33 Moore, C.M. *et al.* (1998) Object-based visual selection: evidence from perceptual completion. *Psychol. Sci.* 9, 104–110
34 Feldman, J. (2002) Perceptual grouping into visual 'objects': a detailed chronology. *Proc.2nd Annu. Conf. Vis. Sci. Soc.*, p. 172
35 Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, Freeman & Co.