Psychological Review
2007, Vol. 114, No. 3, 599–631

Copyright 2007 by the American Psychological Association
0033-295X/07/$12.00 DOI: 10.1037/0033-295X.114.3.599

A Boolean Map Theory of Visual Attention

Liqiang Huang and Harold Pashler
University of California, San Diego

A theory is presented that attempts to answer two questions. What visual contents can an observer consciously access at one moment? Answer: only one feature value (e.g., green) per dimension, but those feature values can be associated (as a group) with multiple spatially precise locations (comprising a single *labeled Boolean map*). How can an observer voluntarily select what to access? Answer: in one of two ways: (a) by selecting one feature value in one dimension (e.g., selecting the color red) or (b) by iteratively combining the output of (a) with a preexisting Boolean map via the Boolean operations of intersection and union. Boolean map theory offers a unified interpretation of a wide variety of visual attention phenomena usually treated in separate literatures. In so doing, it also illuminates the neglected phenomena of attention to structure.

Keywords: visual attention, consciousness, visual feature

Visual attention, in its most fundamental sense, is a selective visual process that governs access to consciousness. Therefore, a theory of visual attention may naturally start with two questions, addressing limitations in *access* and mechanisms of *selection,* respectively. These questions boil down to the following: What can an observer visually consciously access (or as we sometimes say, *apprehend*) at any given moment? And how do the mechanisms of visual selection govern the choice of what is accessed?

The tasks and limitations studied in the great majority of previous work in the visual attention field (e.g., visual search tasks; Duncan & Humphreys, 1989; Treisman & Gelade, 1980; Wolfe, 1994) relate to the process of selection because the key manipulations in these studies have involved varying the quantity and nature of distracting information (e.g., set size manipulations, feature search vs. conjunction search) in situations in which only one single attribute is to be accessed and reported. In such cases, separating the relevant information (target) from irrelevant information (distractors) is the chief difficulty posed by the task. Accessing the to-be-reported feature of the target, on the other hand, is a relatively minor aspect of the task and one whose difficulty is usually held constant across different conditions of a given study.

Attentional limitations in access have been the focus of only a few studies, most notably those of Duncan and colleagues dealing with a phenomenon that he termed *target–target competition* (Duncan, 1980a, 1980b). In Duncan's experiments, the separation of relevant information from irrelevant information was deliberately made very easy; the results provided evidence that conscious apprehension of multiple bits of relevant information was subject to severe capacity limitations not involved in selection per se. The distinction between limitations in access and selection is concretely illustrated in Figure 1. An observer tries to perceive the color(s) of square(s) in brief exposures. The access is easy in the top row (only one color has to be accessed) but difficult in the bottom row (two colors have to be accessed). The selection is easy in the left column (no distractors) but difficult in the right column (many distractors). On the one hand, most visual search studies basically examine the situation depicted in the top right corner. On the other hand, Duncan's (1980a, 1980b) studies pertained to the bottom left corner.[1]

The literature on visual attention has made significant progress in characterizing the principles of selection, based upon 25 years of active research on visual search, much of it sparked by the pioneering work of Treisman and Gelade (1980; see Quinlan, 2003, for a recent review). On the other hand, limitations of access have been studied much less extensively, and here, findings have basically been restricted to showing that access is subject to severe capacity limitations (e.g., Bundesen, 1990; Duncan, 1980a, 1980b).

To see the sort of issues that remain unresolved in the field, consider Figure 2. What can a person consciously apprehend at any given instant when he or she tries to take in all the information

Liqiang Huang and Harold Pashler, Department of Psychology, University of California, San Diego.

This research was supported by National Institute of Mental Health Grant R01-MH45584 to Harold Pashler. Special thanks to Anne Treisman, whose comments and prior writings were of enormous assistance in helping us to formulate and present the theory described here. We also thank James Enns, Patrick Cavanagh, Jeremy Wolfe, Kyle Cave, Donald I. A. MacLeod, William Prinzmetal, Nancy Kanwisher, Steve Link, Alex Holcombe, and Edward Vul for very helpful comments and/or discussion.

Correspondence concerning this article should be addressed to Liqiang Huang, who is now at the Center for the Study of Brain, Mind and Behavior, Princeton University, Green Hall, Princeton, NJ 08544. E-mail: lhuang@princeton.edu

---

[1] This selection–access distinction is somehow similar to the concepts of capacity–selectivity widely used in the literature (e.g., Desimone & Duncan, 1995; Pashler, 1998) where capacity is related to limitations in access and selectivity is related to limitations in selection. However, the two have often been confounded; for example, in a very difficult visual search task, the challenge should be selectivity, but researchers often suggest that attentional capacity is limited in such tasks.

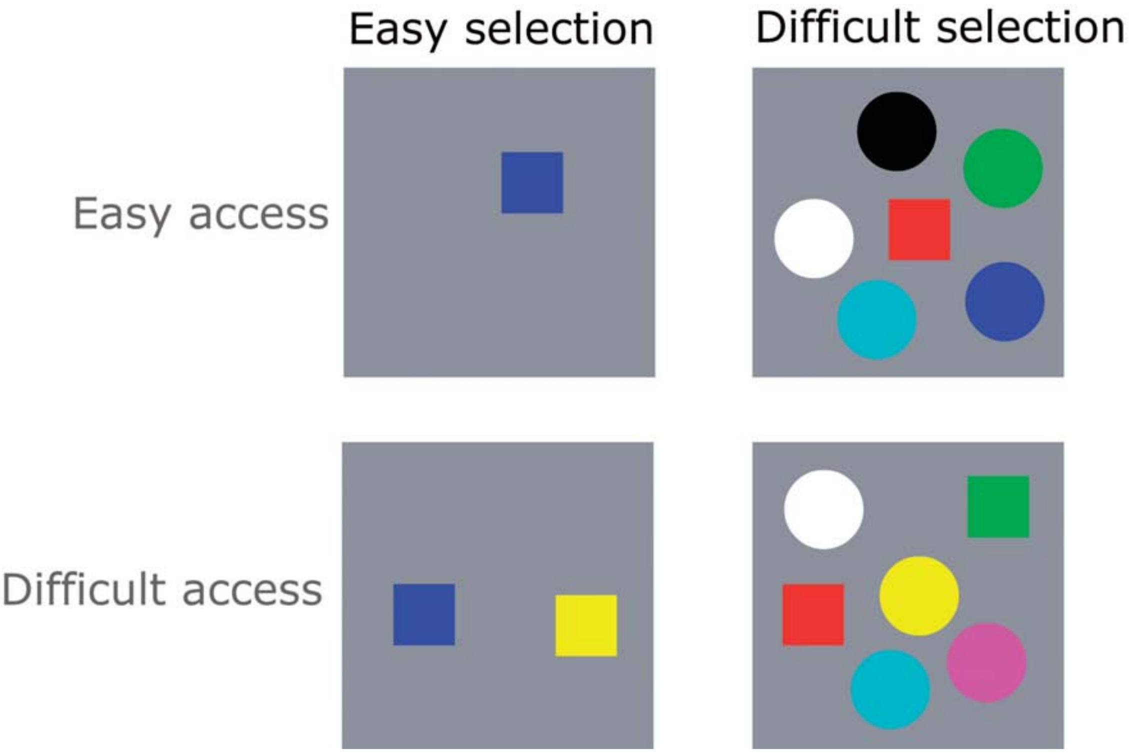


Figure 1. Various displays with which an observer can carry out the task of perceiving the color of the squares. Two types of attentional limitations (selection, access) are illustrated. Columns indicate whether selection of the relevant elements (squares) is easy or difficult (based on presence of disks as distractors). Rows indicate whether access to the relevant information is easy or difficult (based on complexity of to-be-accessed information).

within each of the panels of this figure? The answer is not obvious on the basis of existing theory and research. Feature-integration theory (Treisman & Gelade, 1980) might seem to speak to this question. After all, the theory states that features are processed in parallel and are simultaneously available, whereas attention is necessary to bind features from different dimensions together into one object. From that, one might infer that access to two features (e.g., two colors) can be achieved in parallel as long as they do not

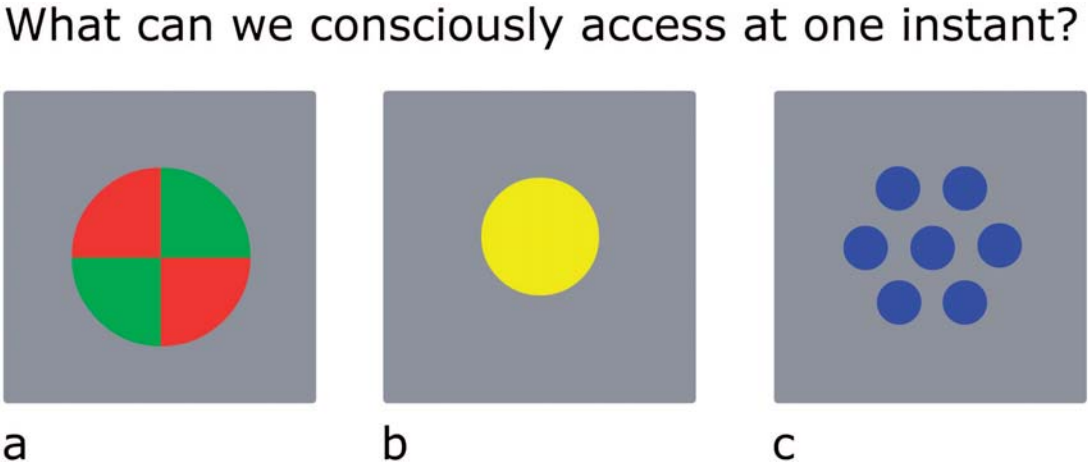


Figure 2. What information can be consciously accessed at any given instant when an observer attends to each of these three different panels? Boolean map theory claims that only one feature (e.g., one color) can be consciously accessed at one instant; therefore, it contends that the configuration of colors and respective locations in Panels b and c can be apprehended at a glance (because only one color is mapped to one or more locations), whereas the configuration in Panel a cannot be momentarily apprehended (because the two colors have to be apprehended one by one).

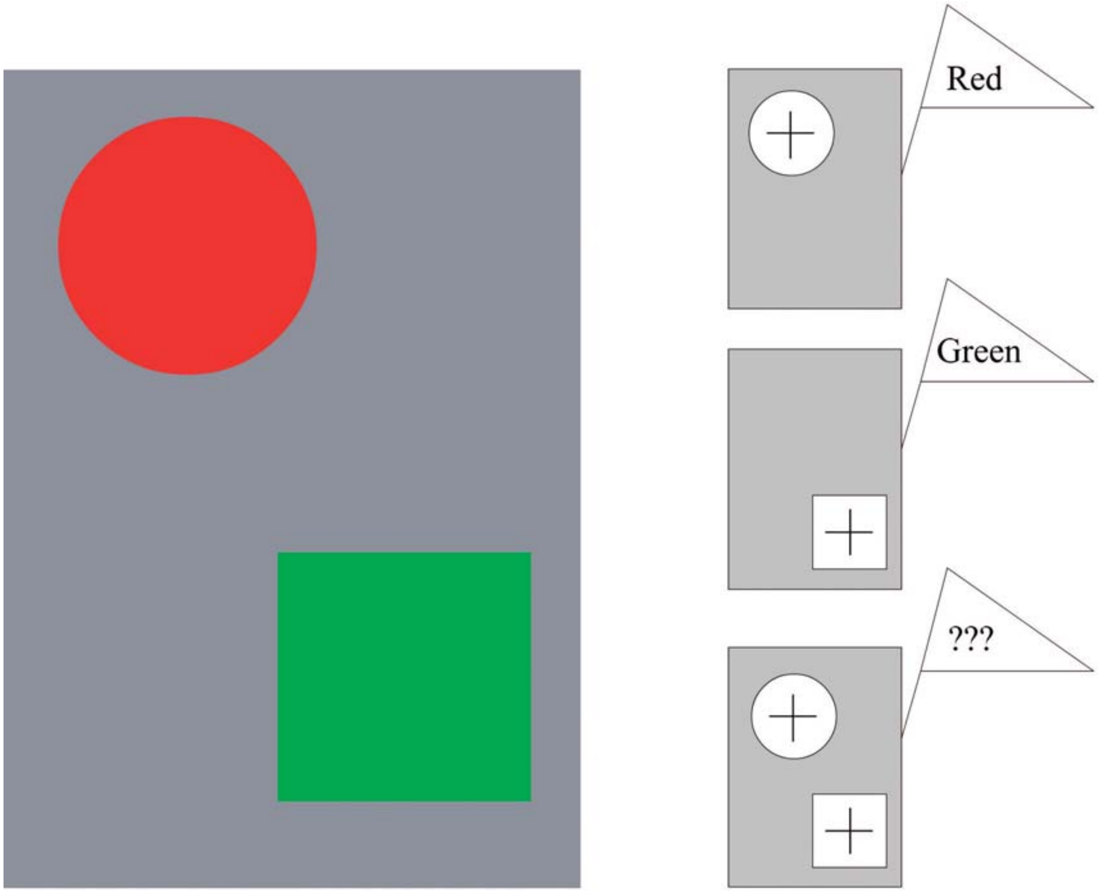


Figure 3. Three possible labeled Boolean maps that could describe the display on the left (composed of a red disk and a green square). One map includes only the disk and indicates its shape and color (a red disk); another includes only the square and indicates its shape and color (a green square). The third map includes both the square and the disk and describes only the global shape, but neither of the two colors.

have to be bound with other feature dimensions. This conclusion is demonstrably wrong, as we show below. At the risk of somewhat oversimplifying the theory to be presented here, the answer to this question offered in the present article essentially boils down to this: What can be accessed at any given instant is only one feature value per dimension, with all these feature values (in their entirety) being perceived to occupy one or multiple locations.

A clear appreciation of the distinction between selection and access is crucial to the present theory because they are typically not formally distinguished in the literature. For example, a question such as "Can we attend to two colors simultaneously?" can mean two entirely different things. When referring to access, it amounts to the question "Is it possible to detect (or perceive) two colors simultaneously?" When referring to selection, it is tantamount to asking, "Can one selection process be based on two colors at the same time?" The answers to these questions are potentially independent in the sense that one can envision theories that would embrace any of the four possible pairs of answers to these two binary questions. Moreover, as we discuss below, the notion of seeing a color can be analyzed in a number of different ways.

A Boolean Map Theory of Visual Attention

What Is a Boolean Map?

Before describing the Boolean map theory in a concrete fashion, we must explain what a Boolean map is. As hinted at above, it is a spatial representation that partitions a visual scene into two distinct and complementary regions: the region that is selected and the region that is not selected.[2] In Figure 3, if a Boolean map is

---

[2] In the present theoretical framework, a Boolean map is a collection of locations. It should be emphasized that those locations are regions exactly covered by relevant stimuli (i.e., one should imagine the map as being, as it were, shrink-wrapped to conform tightly to the object). For example, in Figure 3, a Boolean map of both objects covers precise regions, and thus, the shape information (square, ball) is also automatically included. This differs substantially from previous theories (e.g., feature-integration theory) in which location seems to be regarded as orthogonal to the form or shape of the object occupying the location (e.g., square, ball, as in Figure 3). Therefore we suspect the word *location* in previous theories has been taken to imply one single coordinate value describing the location of something like the centroid of the entire object.

created to represent only the red object or only the green object (two natural possibilities, but not the only ones), then the other object would be missing from the selected region of the Boolean map. If both are selected, then as far as the Boolean map is concerned, they become indistinguishable: No differentiation in the featural properties of the objects is possible with respect to color, shape, or any more abstract property. The word *Boolean* refers to the fact that the visual scene is divided into only two binary (i.e., Boolean) levels.

The Boolean map may be associated with what we call *featural labels*. It is a key claim of the theory that the Boolean map as a whole may have only a single featural label per dimension, and this featural label must provide an overall featural description of the entire region. That is, there is no mechanism for associating a featural property with any subset of the Boolean map smaller than the whole. For example, when a Boolean map of the display shown in Figure 3 encompasses both the red object and the green object, there may be a color label, but if so, this color label cannot provide access to redness or greenness. To access the redness, one would need to create a Boolean map that covers only the red object. However, there can be independent featural labels for different dimensions of a single map. For example, a single Boolean map could have redness as a color label and verticalness as an orientation label. For the sake of simplicity, this principle of one feature value per dimension at any instant is hereafter often referred to as one feature value at an instant, as a shorthand for the proposition that there can be a separate feature label (but only one) for each dimension and that the label cannot be associated with anything less than the Boolean map in its entirety.

Also, at the risk of belaboring what may already be obvious, we should point out the Boolean map is a format in which one feature value is tied to potentially multiple locations. These locations are encoded by the pattern of Boolean map itself, so, unlike other feature values, multiple locations can be represented simultaneously by one Boolean map.

Boolean Map Is the Mechanism of Visual Access

The preceding paragraph describes the Boolean map as an abstract data structure. Naturally, we introduce the concept of Boolean map to argue that it corresponds to a real and fundamental aspect of human vision. The labeled Boolean map, it is argued here, corresponds to the information that can be consciously apprehended at one instant based on vision. To put it another way, an observer's visual awareness corresponds to one and only one Boolean map at any given instant. This momentary conscious apprehension provides access to both the shape of the Boolean map (the set of potentially multiple location values) and the identity of associated feature labels and represents the latter as properties belonging to the former. Thus, to apprehend the redness of a red object, one must create a Boolean map that covers the red object. We contend that observers are perfectly able to construct a Boolean map that covers both a red object and a green object, but if this is done, neither redness nor greenness is consciously accessible at that instant (as in the bottom right example of Figure 3).

The account described above can be roughly summarized as three tenets for easier reference, though they should not be interpreted as isolated propositions. These are *obligatory encoding of location* (consciously accessed visual information is always indexed by location), *single-feature access* (only one feature value can be accessed at one time), and *multiple-location access* (multiple locations can be accessed at one time).

We have proposed the format of Boolean map as characterizing the content of conscious access at any one instant. This is an attempt to answer the first question about access raised in the beginning of this article (What can an observer visually consciously access at any given moment?). The second question described there, relating to selection (How do the mechanisms of visual selection govern the choice of what is accessed?), requires us to describe the possible ways in which a Boolean map can be created, and this is discussed in the next section. In short, then, Boolean map theory addresses the two questions about access and selection by (a) specifying the data format of a Boolean map and (b) constraining the possible ways of creating a Boolean map, respectively.

Given that feature-integration theory (and its subsequent offshoots) does not explicitly distinguish between selection and access and—at least, read narrowly—is restricted to characterizing selection, it is not a simple matter to relate these propositions to that tradition. A more directly relevant line of work is found in Duncan (1980a, 1980b), who—as already mentioned—documented the difficulty of simultaneously accessing two targets even when the selection process is easy. One of the ways that the theory presented here goes beyond Duncan's analysis is in proposing that even though observers cannot simultaneously access two feature values, they can simultaneously access two or more location values.

Creation of Boolean Map

A second postulate of the present theory is that all top-down control over visual perception is accomplished by processes that trigger and guide the creation of a Boolean map. This creation is envisioned as occurring through one of two different possible routes. The first route is creating a Boolean map de novo by selecting one feature value from a single dimension (which could include location itself, through processes described below). Here, the location or locations that characterize the distribution of that feature value are included in the resulting Boolean map (e.g., selecting all red objects in a visual scene results in a labeled Boolean map extending over the region comprising red things). The second route is through combining the preexisting Boolean map with the output of the first route (selection based on one feature value) via the Boolean operations of intersection and union—subject to important limitations described below. The theory also makes the strong claim that these two routes exhaust the mechanisms of selecting one set of visual information or another and routing it to consciousness. From that, two additional important conclusions follow. First, only one feature value can be used to create a Boolean map at one time. Second, a Boolean map cannot be created by simultaneously combining information from more than one dimension (i.e., creation is always based on one dimension at one time). These claims obviously run counter to a significant amount of opinion in the field, and the reader may find their import rather vague and abstract at this point. Each of them will be elaborated upon, rendered more concrete, and defended below. To avoid misunderstandings, it should be noted that the Boolean operations postulated in our discussion of the mechanisms of Boolean operations are fully consistent with the claim that at any given instant there is only one Boolean map because new Boolean maps created through Boolean operations automatically replace any preexisting map. The claims above can be roughly summarized as two tenets. *Feature-by-feature selection* (selection can be based on only one feature value at one time) and *availability of Boolean operations* (any selection that cannot be directly accomplished by selection based on one feature value must be created through Boolean operations).

It should be noted that the Boolean map is a relatively spatially precise and potentially complex spatial representation that can encompass spatially disconnected regions. This is of course natural given that the Boolean map can be created by selecting a feature, and the distribution of features can obviously occupy disconnected regions within a display. Thus, the Boolean map theory implies that terms like *spotlight* or *zoom lens* are entirely inadequate—even as metaphors—to describe the character of visual attention (e.g., Driver & Baylis, 1989; see also LaBerge & Brown, 1989).

At the risk of stating what may now be obvious, a feature value used in governing the creation of a Boolean map can be voluntarily chosen from all the feature dimensions present in a particular display. For example, in a display in which objects vary in several color values and several orientation values, the Boolean map can be voluntarily created either by selecting a color or by selecting an orientation. It is important to note, however, that a Boolean map can also be created by selecting all the objects in a display. We have already seen an example of this in the selection of both objects in the lowermost Boolean map depicted in Figure 3.

Summary of the Boolean Map Theory

According to the theory examined here, there are two different kinds of visual attentional limitations. One is defined by constraints on the available ways to create a Boolean map (i.e., selection). The other is defined by the information that can be contained in a Boolean map (i.e., access). Describing these constraints will begin to flesh out our proposed answers to the two questions raised earlier: What can an observer visually consciously access at one moment, and how do observers select what to access?

Thus, Boolean map theory can be briefly encapsulated in two propositions.

1. *Principle of access.* A Boolean map is a spatial representation that divides the visual field into two discrete subsets: selected and not selected. This map can be associated with featural labels, but no more than one per dimension. A feature label must provide a single overall featural description of the entire region encompassed by the map. At any given instant, an observer's conscious awareness of a visual scene can be represented by a single labeled Boolean map.

2. *Principle of selection.* All top-down control in visual perception is accomplished by directing the ways in which a Boolean map is created. This takes place in one of two ways: (a) by selecting one feature value in one dimension or (b) by combining the existing Boolean map with output of (a) through one of two Boolean operations (intersection and union).

To take it further, the essential points of Boolean map theory can be further condensed into the five tenets we have given above: obligatory encoding of location, single-feature access, multiple-location access, feature-by-feature selection, and availability of Boolean operations.

The above presentation of the theory seeks to characterize what is basically a data format underlying conscious access and the possible ways of generating that data format. The theory may also be viewed as a description of how selection can be undertaken and what results emerge from the process of selection. Though the substance is the same, the latter approach may strike some readers as more familiar. Again, the same five tenets figure in this more narrative mode of describing the theory, which runs as follows: The visual system can select one feature at one time (feature-by-feature selection), and the process yields a map of the spatial distribution of the feature (obligatory encoding of location) with the option of combining this map with a preexisting map (Boolean operations). In each selection process, only one feature value can be accessed from one dimension (single-feature access), but the map itself provides access to multiple locations (multiple-location access).

Figure 4 shows a schematic representation of Boolean map theory. Here, visual processing is divided into two stages: feature maps and Boolean map. Feature maps are generated in early vision (the two subroutines relating to feature maps—feature-location routine and location-feature routine—are described below). The Boolean map is the sole mechanism of conscious access, with creation of Boolean maps being the sole means of top-down control.

Some Additional Comments on Concepts Introduced Here

Feature maps. In the preceding discussion, we have used the terms *feature* and *feature map.* In using these concepts, we are of course borrowing ideas common throughout much of the vision literature. This assumes the existence of a set of mechanisms that can extract featural descriptions from visual input without the generation of a Boolean map (Treisman & Gelade, 1980). Presumably, the feature maps include those computing motion, size, color, spatial frequency, orientation, and perhaps others (Treisman & Gormican, 1988). However, it follows from what has been said already that those features cannot be consciously accessed until they are represented as labels on a Boolean map and that they participate in generating conscious experience only according to the two rules of Boolean map formation stated above. Also, to clarify terminology, in this article, the word *feature* is used to refer to feature values (e.g., red and green are features), whereas color and orientation are always referred to as *dimensions.*

There are two natural ways of using visual information in the feature maps, which we describe as two subroutines. The first takes as an input a featural value and returns a Boolean map describing all the locations at which that feature value is present (this is called the *feature-location routine* and abbreviated *FL*). This *FL* is somewhat similar to prior concepts such as feature-based attention or the guidance process in Wolfe's guided search model (Kim & Cave, 1995; Wolfe, 1994) or Hoffman's two-stage model (Hoffman, 1979). The second subroutine takes as input a location value and returns a featural value for that location (the

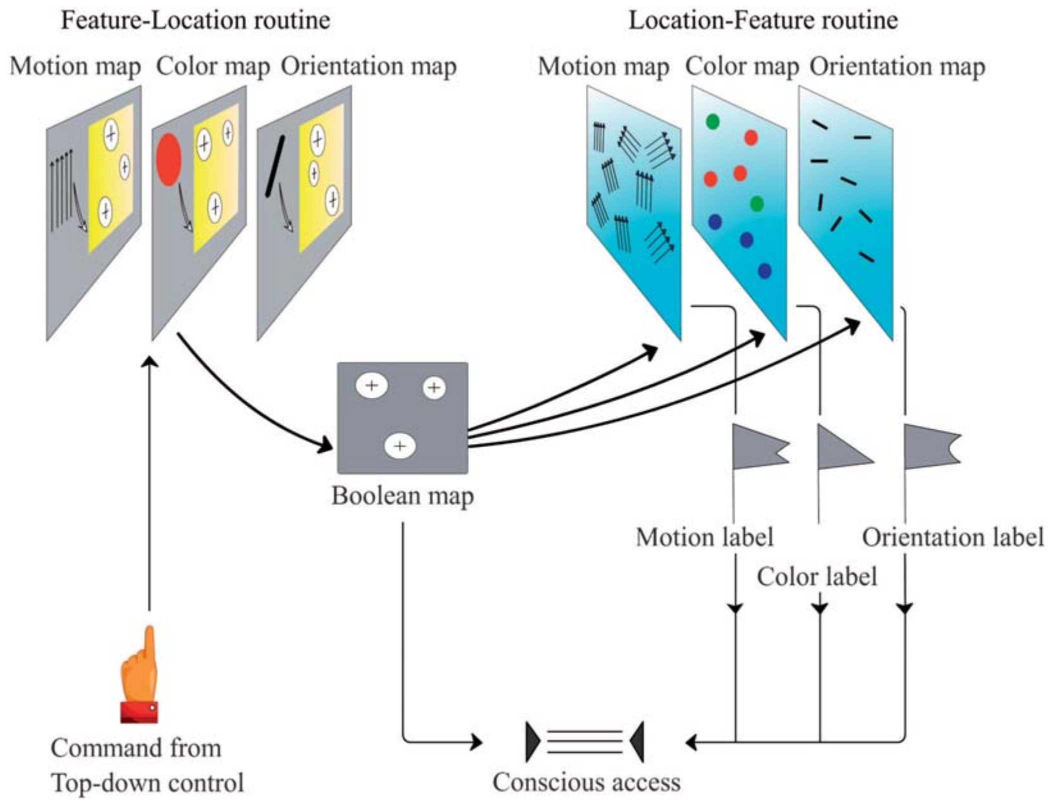


Figure 4. A schematic overview of Boolean map theory. The sensory analysis begins at the top through creation of the feature maps. Top-down control is (solely) implemented by giving commands to the feature-location routine to trigger creation of a Boolean map. The Boolean map (along with the feature labels sent back by the location-feature routine) constitutes the visual information that is consciously accessible at one moment.

location-feature routine, abbreviated *LF*).[3] In Boolean map theory, the *FL* is responsible for creating the Boolean map according to a top-down control value (e.g., seeing everything red, seeing the object on the top left corner). The *LF* is responsible for creating labels from all the dimensions (i.e., allowing them to reach conscious access) once a Boolean map is created.

A formal representation of Boolean map theory. Here, we give a very simplified formal representation incorporating the basic aspects of Boolean map theory to help avoid the potential ambiguity present in any verbal presentation, such as the one offered above.

Definitions: L is the location variable. X and Y are two feature variables. $L = l_1$ means the location variable is given the value l_1. $X = x_1$ means Feature X is given the value x_1. The arrow symbol ($\rightarrow$) means *returning the value.*

So, the basic sequence of events in attentional selection of visual information is as follows:

1. Specifying one control value voluntarily, for example, (x_{control}) or (l_{control}). This value is used in *FL* below.
2. From this control value, a set of locations is retrieved from *FL*: $FL(x_{\text{control1}}) \rightarrow (l_1, l_2, l_3, \ldots)$, or $FL(l_{\text{control1}}) \rightarrow (l_1, l_2, l_3, \ldots)$.
3. The set of locations could be combined with the locations of the preexisting Boolean map through Boolean operations.
4. For each of the locations values ($l_1, l_2, l_3, \ldots$), a set of feature values is retrieved from *LF*: $LF(l_1) \rightarrow [x(l_1), y(l_1)]$; $LF(l_2) \rightarrow [x(l_2), y(l_2)] \ldots$.
5. Those feature values are used to compute the feature labels: $[(x(l_1), x(l_2), \ldots] \rightarrow x_{\text{label}}$.

---

[3] The effectiveness of these two subroutines plausibly depends on how information in the feature maps is organized. Only the *LF* would be possible in a map that represented only the feature value in each location, which seems to have been the conventional (though usually only implicit) conceptualization of a feature map in previous theorizing. By way of analogy, consider a telephone book that lists 100,000 names (corresponding to locations) in alphabetical order, providing a telephone number for each (corresponding to a featural value). This representation will support something akin to the *LF*: rapidly finding the telephone number (feature value) for each desired name (location). However, as everyone knows, it is difficult to find the name that goes with a particular telephone number, that is, an ordinary phone book does not support the *FL*. To effectively find a name for a particular number, one needs another book (a reverse index) that lists all the names in the order of telephone numbers. The *FL* plausibly depends on such a reverse index.

When those feature values are all equal (i.e., selected region is homogeneous in that dimension), the label has an explicit value.[4] If $x(l_1) = x_1$ and $x(l_2) = x_1 \ldots$, then $x_{\text{label}} = x_1$.

Thus, the theory presented in this article can be summarized as follows:

1. *Principle of access.* Conscious access has the format of a labeled Boolean map: The set of consciously accessible information can be represented as: (l_1, l_2, l_3, ...; x_{label}, y_{label}). In this format, there can be multiple location values, but only one feature value can be represented for each dimension.

2. *Principle of selection.* The Boolean map is created by selecting one feature value in one dimension at a time (e.g., x_{control} or l_{control}), with the option of Boolean operations with the preexisting Boolean map.

Revisiting the distinction between selection and access. Above, we suggested that although a distinction between selection and access had made a few cameo appearances in the literature on visual attention, the two types of account have largely been conflated in most theoretical writings. This point can be highlighted in a more concrete fashion given the concepts introduced in the previous section. The control value (feature or location) sent to *FL* is distinct from the conscious percept that is returned by *LF*. In both cases, we argue that there can be only one feature value (to be accessed/to be used to select) at one instant, as stated in the tenets of single-feature access and feature-by-feature selection, respectively. Although these two claims may at first sound redundant, they are logically quite distinct. That is, one can easily imagine a hypothetical theory that would maintain only one of them and not the other. For example, a theory might maintain feature-by-feature selection and claim that one can select all vertical objects and then have simultaneous access to all of their three colors. On the other hand, a theory might uphold single-feature access while maintaining that one can immediately perceive the spatial pattern of green and blue objects from a background of red and yellow objects.

Predictions from Boolean map theory. Here, we note a few of the most immediate and important predictions that follow from Boolean map theory. From the claim that conscious access has the format of Boolean map (l_1, l_2, l_3, . . . ; x_{label}, y_{label}), one may infer two things. First, only one feature value can be accessed from each dimension at one instant (single-feature access), whereas multiple location values can be accessed simultaneously (multiple-location access). For example, in Figure 5a, accessing the colors of the balls requires a series of Boolean maps selecting individual balls, but accessing the locations of all the balls (i.e., the spatial configuration or pattern) requires only one Boolean map selecting all balls. Second, access to featural information is always accompanied by awareness of the location of the accessed feature (i.e., if an observer sees something, the observer always knows where it is: obligatory encoding of location). We should point out, in case it is not already self-evident, that both of these predictions are in fundamental disagreement with the most famous aspects of feature-integration theory. According to feature-integration theory, features are processed in parallel and are simultaneously available as long as they do not have to be integrated. Thus, access to multiple features can be achieved in parallel, and these do not have to be accompanied by location information. The first prediction also runs against ordinary intuition, which seems to suggest to those inclined to reflect on the matter that an observer can continuously perceive multiple features at the same time (e.g., in Figure 2a).

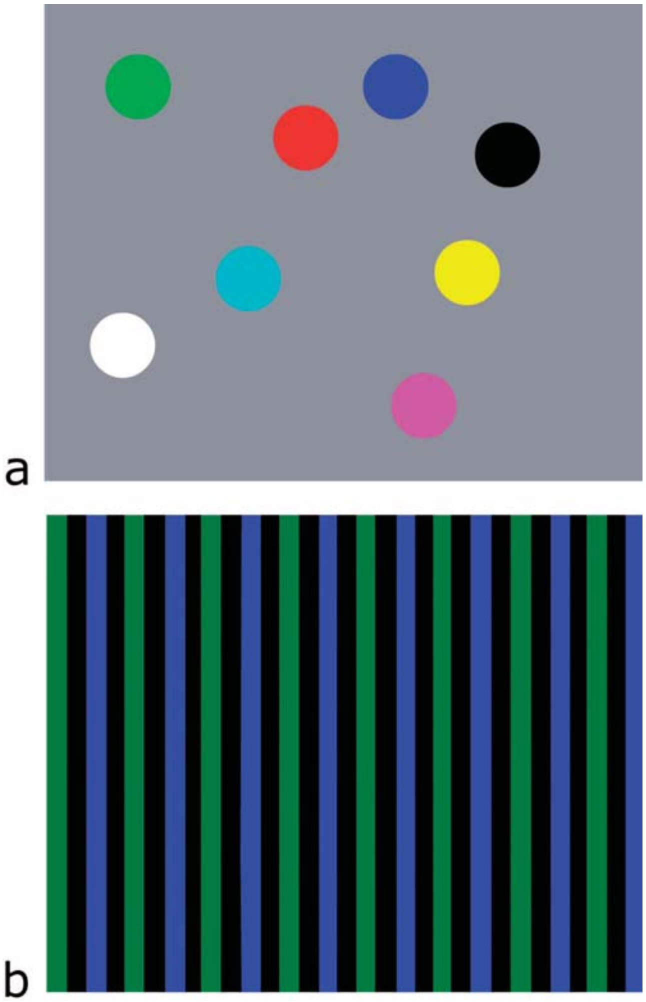


Figure 5. Two demonstrations of the claim that the selected region itself can serve as data for visual analysis. Attending to the red, white, and yellow disks in Panel a, an observer can see their triangular arrangement. In Panel b, attending to the blue lines or the green lines, one sees a periodic pattern; attending to both the blue and the green lines, one sees a periodic pattern with twice the frequency.

The claims about selection described above put two important constraints on the possible operations of selection. First, the feature-by-feature selection tenet rules out the idea of making a selection based on more than one feature value simultaneously (e.g., dividing color space into two parts and selecting one of them;

---

[4] Boolean map theory does not specify what label would be used when the selected region is not homogeneous on one dimension. Even if such a label cannot provide conscious access to individual feature values, it is probably able to provide an overall statistical description of all the feature values of the selected region (Chong & Treisman, 2003, 2005).

Bauer, Jolicoeur, & Cowan, 1996; D'Zmura, 1991). Second, the availability of Boolean operations tenet claims that any selection that cannot be accomplished by selection based on a single feature is implemented by Boolean operations. Thus, it rules out the idea of weighting signals from multiple dimensions into one single scalar according to behavior goal (e.g., Wolfe, 1994). Both claims are elaborated upon and defended later in this article.

Attention as structure and data. According to traditional theories of attention, preattentive visual machinery generates many visual encodings, and attention selects, enhances, or inhibits some of these. Selection of a region is thus viewed as a signal that controls processing rather than as data to be analyzed by visual processing routines. By contrast, in the theory suggested here, the spatial pattern of the Boolean map (i.e., the shape of the selected region) is data available to conscious access by itself.[5]

It is not difficult to construct simple displays that demonstrate the potential for such analysis. Consider Figure 5a. If cued to attend to any three balls (e.g., the red, white, and yellow ones), an observer can perceive the triangle with corners located at the balls after the balls have been selected in three steps. The location and shape of the triangle depend on which three colors are picked. Given the number of possible triangles, it seems unlikely that all combinations are analyzed as triangles by preattentive visual machinery prior to the selection. Therefore, perceiving that triangle requires spatial analysis of the pattern of Boolean map itself rather than selection of visual information already computed in early vision. A similar example is given in Figure 5b. An observer can choose to attend to only blue lines or only green lines and see a relatively low spatial frequency pattern. Alternatively, the observer can choose to attend to both blue and green lines and see a pattern of twice the spatial frequency.

Despite the ease with which examples of this kind can be created, they have rarely been discussed in past literature on attention. The first example mentioned above (a triangle from three selected balls) has sometimes been mentioned under the heading of *visual interpolation.* Besides that, the spatial analysis of the selected region has apparently been considered in depth only within a few scattered theoretical treatments: the study of visual routines (Ullman, 1984), attentional tracking (Pylyshyn & Storm, 1988; see also Yantis, 1992, for an explicit approach of shape analysis), motion perception (Lu & Sperling, 1995), and what Cavanagh, Labianca, and Thornton (2001) and colleagues termed *sprites.* In each case, an intriguing phenomenon was noticed, and significant empirical findings were noted. However, the underlying principle of analysis of the selected region has not been elevated to any particularly important place in theorizing about visual attention. The present article contends that new insight into a variety of perceptual and attentional phenomena may be gained if this concept is elevated to a very prominent place.

Moreover, by conceptualizing attention as a spatial pattern that itself provides data, the theoretical account developed here draws attention to a broader range of phenomena than those upon which visual attention theories have traditionally been based. Most research has focused on visual search tasks in which a person scans a complex display with the goal of finding a (usually singular) target. The present account seeks to illuminate a wider array of ways in which attention enables the perception of visual structure. Of course, visual search can be described as a sort of structural analysis (discerning the presence of a target), but the structure detected usually involves only a small aspect of a display. This extension may depart from previous research but not from the intended domain of previous theories; for example, although feature-integration theory has popularized the visual search design, the theory itself has been built upon evidence from converging paradigms including illusory conjunction, texture segregation, object files, and so on and was never intended to be focused specifically on visual search.[6]

Boolean Map Theory Compared With Previous Theories

Many components of Boolean map theory resemble elements in previous theories. Table 1 lists some of the claims and other unique aspects in Boolean map theory and refers to some of the theories that these claims agree with or challenge. This table should help to place the present framework in the context of the literature.

In brief, Boolean map theory can be related to previous theorizing in the following very general way. Duncan (1980a, 1980b) distinguished two types of attentional limit: capacity (limitation of access) and selectivity (limitation of selection). Previous research in the area of visual attention (Treisman & Gelade, 1980; Wolfe, 1994) has made substantial progress in uncovering the principles of selection. However, these and other theories have not offered any qualitative or formal constraints on access—on the information inherent in a person's momentary visual awareness. In Boolean map theory, both access and selection are characterized using the concept of the Boolean map.

Boolean Map as the Mechanism of Visual Access

The previous section described the claims of Boolean map theory largely in the language of data structures because we believe this provides the clearest and most appropriate rubric for the theory. However, the reader probably found this relatively abstract formulation unfamiliar and perhaps even strange. Our goal in the next several sections of this article is to persuade the reader that although this formulation may be unusually abstract, it not only accounts for many concrete but hard-to-reconcile results within the familiar lines of visual attention research but also points to many new kinds of visual phenomena that have not been much explored but deserve exploration—especially those relating to the perception of structure in complex displays. As noted above, the theory proposed here assumes that a Boolean map is the only mechanism through which visual information (featural labels and shape of Boolean map) can be consciously accessed. We first describe evidence consistent with the basic hypothesis that only one feature value per dimension can be accessed from the Boolean map. Then, we show evidence that there can indeed be separate

---

[5] One might suggest that it would be simpler to assume that the shape of the passed information, instead of the pattern of the Boolean map, is being analyzed. For example, in Figure 5a, if a filter removes all the other balls except the three, the scene will appear to be like a triangle, identical to a shape analysis of the Boolean map. Although this point has some merit, a shape computation must be performed to make the three dissociated balls appear as a triangle. As we show below, the feature information is discarded in this shape analysis. So, this interpretation may in the end not be distinguishable from the proposed shape analysis of the Boolean map.

[6] A. Treisman (personal communication, February 10, 2006).

Table 1
Aspects of Boolean Map Theory

Tenets and unique aspects	Previous similar claim	Previous opposite claim
Single-feature access (only one feature value can be consciously accessed at one instant)		Treisman & Gelade (1980)
Obligatory encoding of location (location information is always selected simultaneously with feature values)		Treisman & Gelade (1980)
Spatial pattern of selected region is analyzed	Lu & Sperling (1995), Yantis (1992)	
Early vision automatically computes a set of separate features maps	Treisman & Gelade (1980)	
Feature-by-feature selection (feature-based selection can only be based on one feature value at one time, not a subset in the feature space)		D'Zmura (1991)
A Boolean operation, not top-down salience, is the strategy of conducting conjunction search		Wolfe (1994)
Distinguishing the attentional limit in access and selection	Duncan (1980a, 1980b)	

featural labels for different dimensions associated with a single map.

Only One Feature Value Can Be Accessed at One Instant

As illustrated in Figure 3, given a display containing both a red and a green object, there are three possible Boolean maps that could be constructed. One is a Boolean map that selects only the red object. Another is a Boolean map that selects only the green object. The third is a Boolean map that encompasses both objects. If this third possibility is elected, then according to Boolean map theory, the individual colors cannot be accessed. Therefore, to access the properties of individual objects (e.g., to determine that the display contains both a red and a green ball), one would need to create two distinct Boolean maps in series. When that is not possible, observers will not be able to make any judgment about those individual features.

Structure of multicolor displays. As mentioned above, studies of visual attention have been overwhelmingly focused on visual search and a few other tasks. However, in daily life, people often view their environment without the goal of finding any particular single target object. Rather, they come to be aware of global structural information about a scene or display. Researchers have studied a number of phenomena that relate to the perception of structure, including global/local or Navon figures (large letters made out of small letters; Navon, 1977) and structure from motion (Koenderink & van Doorn, 1991). In this article, we focus on the role of visual attention in spatial transformation tasks (symmetry perception, matching, and mental rotation).

Most previous studies of symmetry perception have focused on symmetry of the spatial configuration of inherently Boolean patterns such as those created by distribution of black dots or lines, as in Figures 6a and 6b (e.g., Barlow & Reeves, 1979; Palmer & Hemenway, 1978; van der Helm & Leeuwenberg, 1996). By contrast, less is known about the perception of symmetry in displays involving multifeature variation, even though these forms of symmetry are evidently important in both nature and art (see Figures 6c and 6d).

The only two studies of symmetry perception that systematically manipulated surface features have come from our own laboratory (Huang & Pashler, 2002; Morales & Pashler, 1999). In the first of these studies, observers were required to judge symmetry in the arrangement of color in a regular grid pattern. That is, the observer decided whether each and every square in the grid had the same color as the corresponding square located equidistant across the axis of symmetry, with asymmetric patterns differing in the color of only one or two squares. The results led to the conclusion (which initially came as a surprise to us) that the symmetry judgment required a serial process in which different subfigures (subsets of the display having a given color, such as red or green) were assessed in series. One finding was that the greater the number of colors, the longer the response times (holding constant the size of the grid). Another, more telling result came from what was termed the ABBA/ABCD method (see Figure 7). Here, we compared responses to two different types of four-color displays in which two squares mismatched. In the ABBA condition, the two mismatched pairs involved only two colors (e.g., one red might be changed to green, and one green changed to red), whereas in the ABCD condition, the two mismatched pairs involved all four colors (e.g., one red changed to yellow, and one green changed to blue). Responses were faster in the ABCD condition, where all four colors were involved. This would be predicted by the hypothesis of sequential scanning through colored subfigures because any mismatch would become evident in the first subfigure checked in the ABCD condition, but not in the ABBA condition.

Morales and Pashler (1999) concluded from their findings that symmetry detection machinery is inherently color-blind. In the present article, this conclusion is endorsed but fundamentally reinterpreted: now, it is contended that an iterative color-by-color Boolean mapping strategy and the underlying poverty of momentary visual awareness that it implies are in no way confined to, or

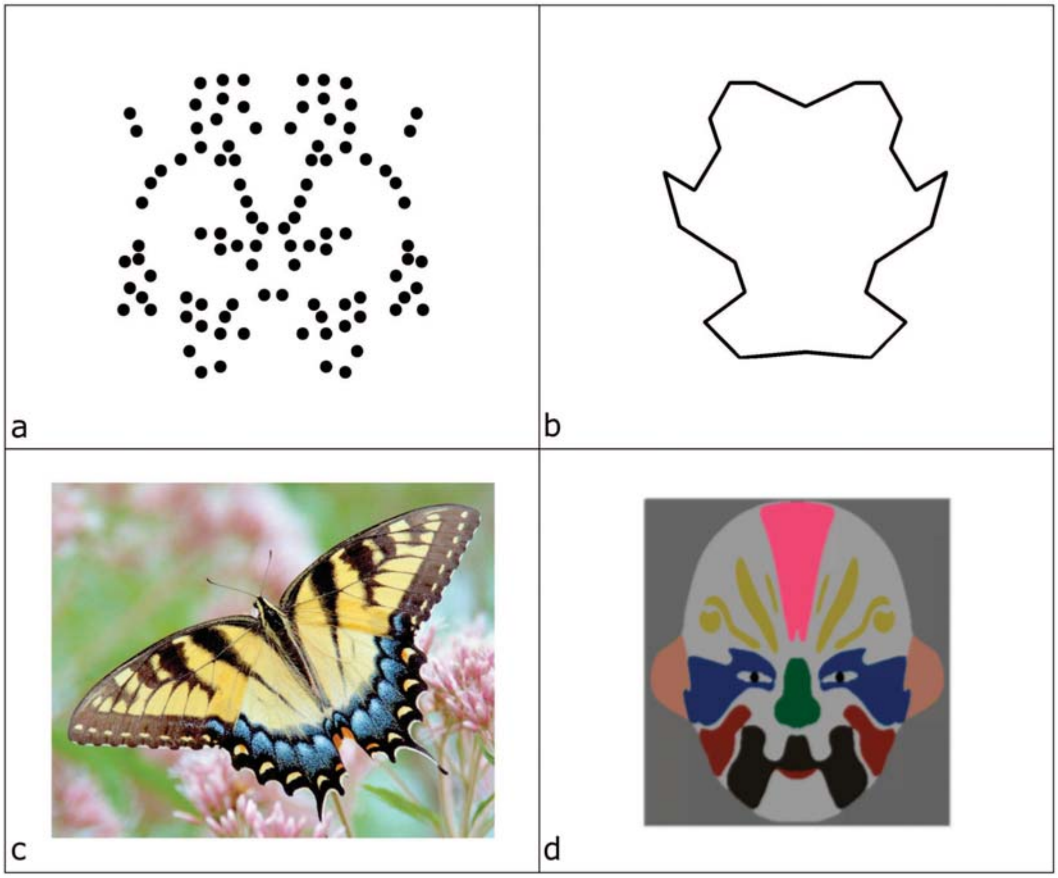


Figure 6. Symmetry perception tasks. Panels a and b show monochrome (and thus inherently Boolean) displays used in most research in this area. Panels c and d show naturalistic color-symmetry displays. Photograph in Figure 6c copyright 2007 by Oscar Gutierrez (www.ogphoto.com). Used with permission.

indeed even particularly related to, symmetry perception per se. Instead, the serial scanning from color to color is the strategy of choice because a Boolean map is the only source of conscious access. Any visual perceptual process that is not already accomplished by early vision would be blind to the spatial arrangement of feature in general and would need to rely upon the selection of only one feature at one time.

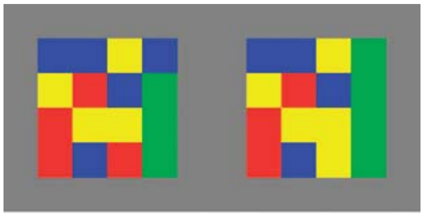
If Boolean map theory is correct, it follows immediately that it should be impossible directly to perceive the spatial structure of a multicolor display, even if an assessment of symmetry is not required. For example, if an observer wants to apprehend the spatial arrangement of the colors in Figure 8, he or she has to construct four Boolean maps in series and carry out a serial scan in which the shape of these maps is assessed.

The first experiments to be described here extend these results from symmetry to the simpler and more fundamental task of pattern matching. Observers judged whether two simultaneously presented displays were identical or not (subject to rotation in one experiment). The details of methods and results are described in the Appendix.

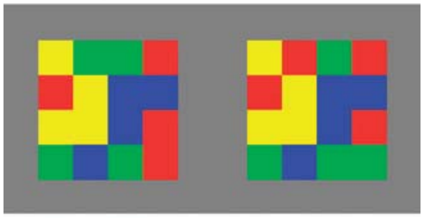
As shown in Figure 8, in the repetition detection experiment (Experiment 1), the observer tried to judge if two color patterns were identical. In the rotation experiment (Experiment 2), the observer tried to judge if two color patterns were identical after 90° rotation. Both experiments used the ABCD–ABBA design from the symmetry studies (Huang & Pashler, 2002; Morales & Pashler, 1999). Performance was indeed significantly better in ABCD displays than ABBA displays in both repetition detection and rotation tasks, suggesting that the same color-to-color scanning principle operated in repetition detection (with or without the additional need for mental rotation).

Figure 9 illustrates the same point more informally. Each pattern is composed of 16 squares, each of a different color. Are the two patterns identical? Unlike in the displays in Figure 8, Boolean mapping should now be highly inefficient (fewer squares per Boolean map), making the comparison exceedingly laborious, as most observers seem to agree that it is.

The single-feature–multiple-locations format. The basic definition of Boolean map describes it as having a single-feature–multiple-locations format. Therefore, it follows that one can only access one feature value (e.g., the color of this spot is green) at one time, but one may simultaneously access multiple locations (e.g., there is a spot here and here and here). Most people seem to agree introspectively that if they have to perceive a multicolor pattern like those in Figure 8, they tend to sequentially shift from subfigure of one color to another. However, viewing a simpler pattern



ABCD Display



ABBA Display

Figure 7. Two different displays used in analyzing color-symmetry perception. In both the ABCD and ABBA displays, the patterns are not color symmetric. However, in the ABCD display, there are two mismatched pairs involving all four colors. In the ABBA display, there are two mismatched pairs, both involving the same two colors. According to Boolean map theory, the ABCD display should elicit quicker rejection because asymmetry is detected in whatever subfigure (red, yellow, green, or blue) happens to be considered first.

such as those in Figure 2a, many people intuit that they can see the two colors simultaneously. We suspect that a sequential strategy still operates[7] but that the sequential steps are less readily introspectable in these cases, perhaps simply because they are faster.

Experiment 3 (see Figure 10a) was conducted to provide an exceptionally direct test of the theory with a method similar to that of Duncan (1980a, 1980b). Here, a two-color pattern was presented simultaneously in one condition and successively in another condition. In both conditions, the stimuli were masked after a very brief exposure. The participants were later shown a test color that could be one of the two colors from the display (half the time) or a completely new color (the other half of the time). If the two colors cannot be accessed at the same time, as Boolean map theory predicts, then the successive condition will have a substantial advantage on the test over the simultaneous condition (Shiffrin & Gardner, 1972). This prediction was confirmed. In addition, a model fitting suggested that the performance difference between simultaneous and successive conditions fit well with the predictions of a strictly sequential model. Experiment 4 (see Figure 10b) tested the issue of accessing two locations simultaneously, also in a successive/simultaneous method. Here, we found that participants did slightly better in the simultaneous condition than in the successive condition. The results of the successive/simultaneous comparison were significantly different in Experiment 3 and in Experiment 4. Taken together, then, these studies are consistent with the idea that people can indeed only access two colors in series, whereas they can access two locations simultaneously. These conclusions follow directly from the theory proposed here, and to our knowledge, they do not follow explicitly from any other theory in the literature.

Is it a problem for Boolean map theory that, when viewing a display like that of Figure 3 (containing a red disk and a green square), many observers claim to have a vivid subjective experience of seeing the two colors at the same time? Before assuming that this is a problem, one should note that the theory does not imply that what such an observer would experience would match the experience of viewing always just one object or of seeing two objects that are gray or some other uniform color. The Boolean map shown in the bottom of Figure 3 implies an awareness of two objects (and perhaps the presence of heterogeneity of color)—what the viewer is claimed to lack is the simultaneous awareness of what colors are present and how they are bound to the shapes. However, that information can be immediately accessed by creating one of the two other Boolean maps shown in the figure. As Dennett (1991) has pointed out in discussing the fact that people tend to be oblivious to the blurriness of their peripheral vision, observers may be unable to distinguish between actually having certain information explicitly represented in awareness and having the ability to access that information quickly whenever they want it (see also O'Regan & Noe, 2001). Thus, if observers can rapidly access the redness of the disk and the greenness of the square in Figure 3 whenever they want to, as the current theory implies, that may suffice to make the observers report that they see both colors at the same time. Additionally, it should be noted that although the present theory challenges some commonsense ways of understanding conscious experience, experimental findings in this area probably make such challenges inevitable. The Boolean map theory proposed here seems to be in less stark conflict with ordinary introspections than are recent views of visual awareness prompted by findings from change detection and change blindness research (e.g., O'Regan & Noe, 2001; Rensink, 2000a, 2000b), a point that we return to below.

So far, we have discussed the issue of accessing two colors or locations in the sense of obtaining information about an individual element. Another, perhaps more important, perspective is that simultaneous accessing of multiple locations makes them a *holistic pattern* and no longer a mere collection of individual locations. A holistic pattern represents associations between locations in the (x, y) plane, reflecting their simultaneous presence in the display. This stands in stark contrast with the fate of features, where no such association can be represented. For example, colors are not organized into a pattern in the color space and thus must be perceived individually. This can be seen in Figure 11. In Figure 11a, one can easily verify that the spatial pattern of disks on the left has the

---

[7] This is consistent with the recent findings of Howard and Holcombe (2007). Howard and Holcombe asked observers to keep monitoring the features or locations of a few objects that were continuously changing. Observers showed a lag in reporting spatial frequencies (they tended to report an "old" feature value rather than the latest feature value), and the lag increased monotonically with number of objects, as expected if the observers serially visited each object and could only report "cached content" from their last visit. More interestingly, no large lag increase was found for reporting locations, consistent with the central claim of Boolean map theory: One must access multiple feature values serially but can have simultaneous access to multiple locations.

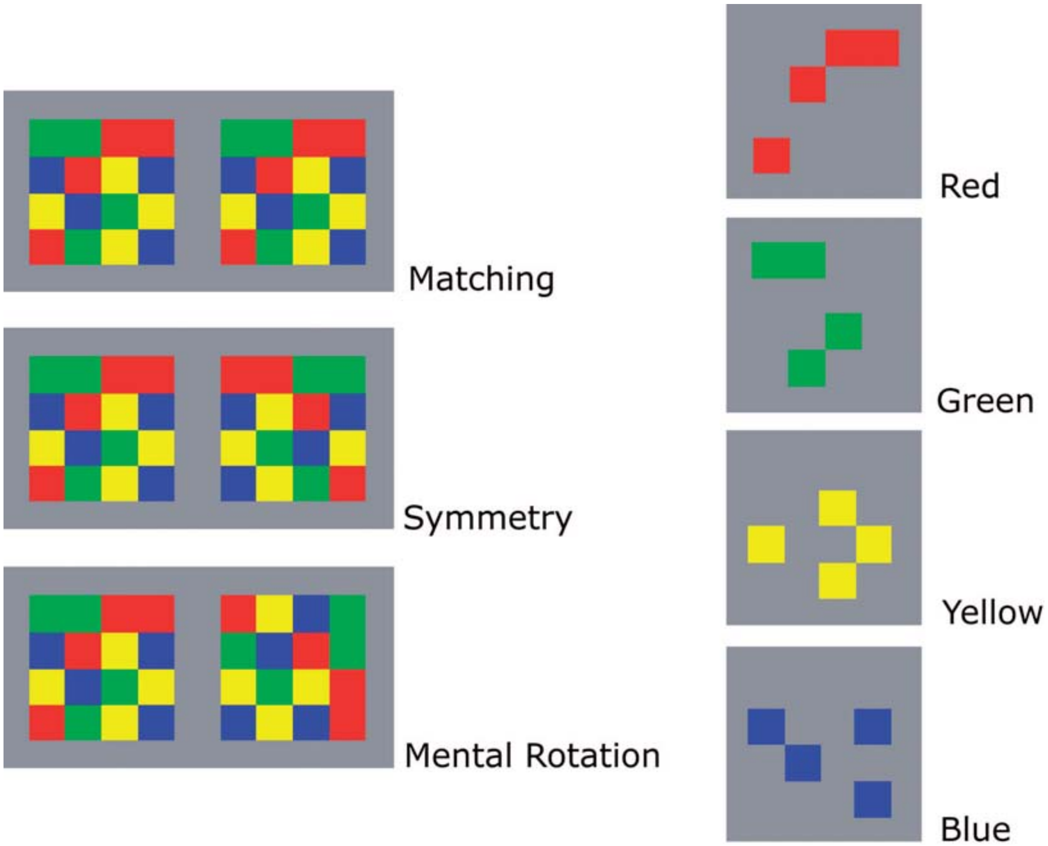


Figure 8. Top left panel: matching task (“Are the two multicolor displays identical?”). Middle left panel: symmetry judgment (“Are the two multicolor displays mirror reflections of each other?”). Bottom left panel: mental rotation task (“Can the left display be rotated to become identical to the right display?”). Boolean map theory claims that such complex multicolored displays must be decomposed into colored subfigures (right panels) before structural analysis can be performed.

same shape as the pattern of disks on the right.[8] By contrast, to verify that the same set of colors is present in the two, one must rely upon a sequential checking (iteratively forming a Boolean map that spans the two displays, one consisting of a pair of yellow items, one of a pair of red items, and so on). The difference between location and other features is even more obvious with respect to their dependency upon each other. Shuffling the disks causes little harm to the comparison of spatial pattern but renders it substantially more difficult to verify the correspondence of colors (see Figure 11b). This makes sense if multiple locations can be perceived together as a holistic pattern, whereas features can only be perceived individually through their locations (see D. G. Watson, Maylor, & Bruce, 2005, for evidence that a sequential checking strategy is used in determining number of colors present in a display). Although this observation may seem intuitively obvious once pointed out, it is not clear to us how prior theoretical analyses would in any way entail it.

Depth is evidently an exception to the single-feature-access tenet. As shown in Figure 12, one can see multiple depth levels at one time, and they form a vivid pattern in three-dimensional space. As mentioned above, multiple locations can be accessed simultaneously. Given that depth is an aspect of location, this exception does not undermine but rather extends and reinforces the plausibility of the present analysis (while offering hints about what coordinate structure Boolean maps are likely to reside in).

Attentional tracking. Attentional tracking has been widely used to study visual attentional limits in general and object-based attention in particular. In this experimental paradigm, a large number of identical objects move randomly in the display. Some of them are highlighted at the beginning of the trial, and participants try to track them while they move. At the end of the trial, participants report identifying properties of the tracked objects. The most fundamental finding about attentional tracking is that participants can track about 4–5 items quite well (Pylyshyn & Storm, 1988).

---

[8] Regarding Figure 11a, we have already, with Figure 5a, talked about the one Boolean map selecting all balls to access their spatial pattern (e.g., all locations). This is not inconsistent with the principle of selection that claims one can only select one feature value at one time. Presumably, in this case, selection is not based on color at all. Instead, it is done by sending a command like “all objects” to the *LF* routine, and the Boolean map created can provide access to its spatial pattern (but not access to color values of balls). This is natural and probably depends on the bottom-up salience that we discuss later.

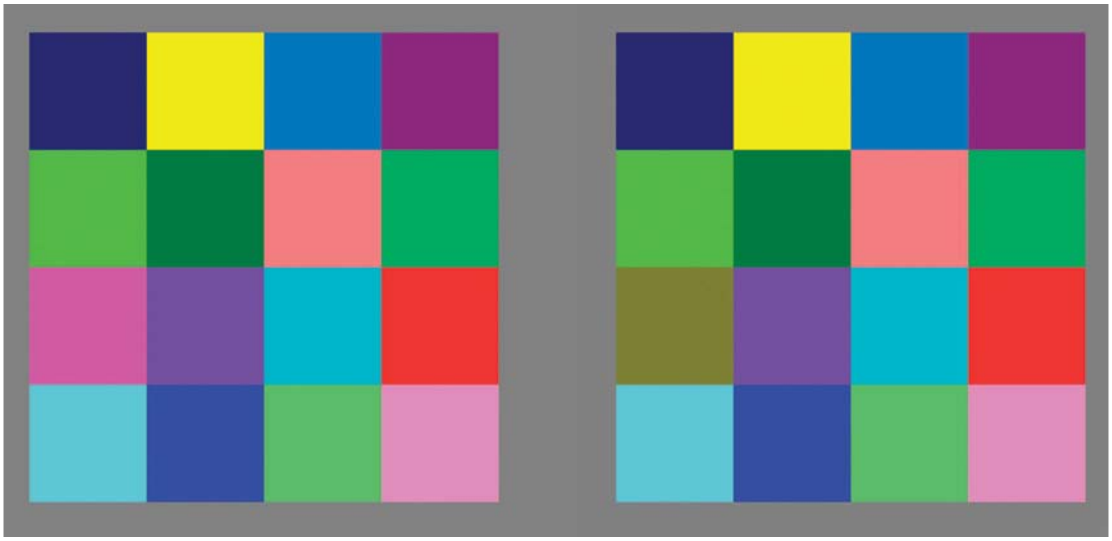


Figure 9. Are the two displays identical? With 16 squares and 16 colors, the task is extremely laborious, according to Boolean map theory, because it cannot be efficiently decomposed into a small number of subfigures of a given color.

Attentional tracking provides a strong test for Boolean map theory because attentional tracking would appear to require the direct use of a Boolean map. In most visual attention studies, the task-relevant information is usually some property of the object, whereas in attentional tracking, the spatial distribution of attention plays a critical role in the task. Also, with static displays, the strict limit of one Boolean map at one time may be masked by the strategy of rapidly switching between one Boolean map and another, whereas in attentional tracking, objects are rapidly moving, presumably foiling any such strategy. A number of important studies from the tracking literature fit very well with the analysis of tracking offered by Boolean map theory.[9]

The most widely discussed account of tracking is the theory of indexing proposed by Pylyshyn (1989). According to this view, each object is listed and differentiated by a pointer, allowing even physically identical items to be treated as different objects. Boolean map theory offers a very different analysis: Because the nature of visual encoding is Boolean, objects are either selected or not selected. Therefore, observers should not be able to maintain any cognitive differentiation between two selected objects during the time all the objects are being tracked. Suppose each ball starts with a different digit in it, and the participants track four balls numbered 1–4. At some point, the digit is erased, and they keep tracking for some time after that. In the process of tracking, participants should readily be able to report which four balls were initially cued. However, they should be unable to report the identities of the objects (i.e., which is which). Note that this prediction goes against indexing theories (Pylyshyn, 1989; see also Kahneman & Treisman, 1984; Kahneman, Treisman, & Gibbs, 1992). However, it was recently confirmed by Pylyshyn (2004).

In addition, the Boolean map theory predicts that participants will be poorly aware of any features of the objects besides their location. For example, when participants track 4 among 10 balls, the Boolean map will not encode the colors or other features of those balls. Two recent studies confirmed this prediction. Scholl, Pylyshyn, and Franconeri (2007) asked participants to track four items. When participants were tracking, one feature of an item was suddenly obscured, and the observer was asked to report that feature. Not surprisingly, participants' memory about location was much better for tracked items than nontracked items. Remarkably, however, there was no difference between the tracked items and other items in the accuracy of reports on color and shape. Similarly, Saiki (2003) asked participants to keep track of three items that rotated in a predictable way: These three balls were evenly distributed in a circle and moved with equal angular velocity. Naturally, this is a very easy task. Participants were asked to report any sudden switch of color between two items. Performance was found to be very poor, suggesting that even if an observer is tracking only three items that rotate in a predictable way, the observer is hardly aware of their features.

Shape analysis of adjacent regions. Boolean map theory also makes strong predictions about the extraction of spatial information from two adjacent regions. As shown in Figure 13, when observers view an adjoining pair of regions, one red and one green, Boolean map theory predicts that the only useful option is to select one region or the other and process the shape of this region; if an observer selects both (which is also an option), then the difference between the two regions disappears, and no access is provided to shape information that depends upon the contour between them.

Most readers will immediately realize that what is discussed here is not a new phenomenon but rather the familiar phenomenon of figure–ground segmentation, which has been studied for a very long time (e.g., Rubin, 1921/2001). Given this literature, there is no need here to document the robustness of this phenomenon. However, there are some interesting points of contrast between the

---

[9] Because we were not aware of these studies when we derived the predictions from Boolean map theory, we view these studies as providing especially strong support for the theory; the reader may or may not agree.

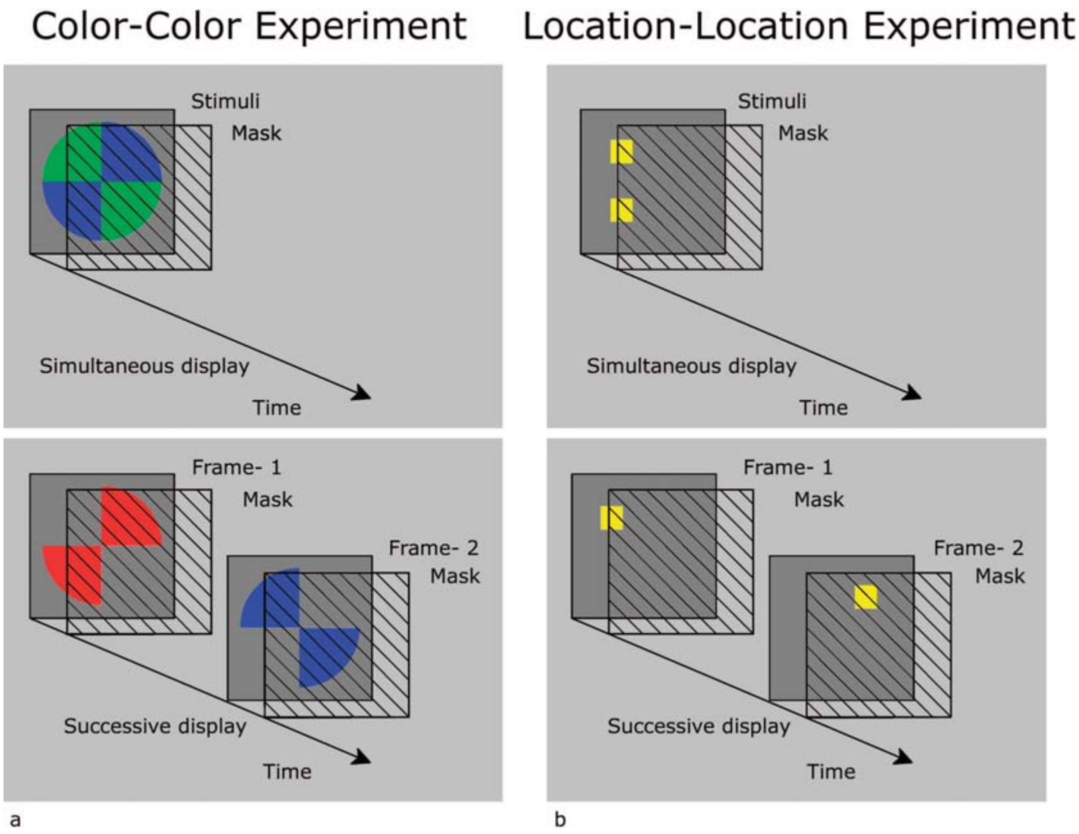


Figure 10. Method of Experiments 3–4. In Experiment 3 (a: color–color experiment), two colors are presented either simultaneously or successively. The participants later decide whether a probe color (not shown here) was in the display or not. In Experiment 4 (b: location–location experiment), two locations are presented either simultaneously or successively. The participants later decide whether a probe square (not shown here) was in one of the locations shown or not. According to Boolean map theory, two features (e.g., colors) can only be accessed sequentially, predicting an advantage for successive displays in Experiment 3. On the other hand, two locations can be accessed in parallel, predicting no advantage for successive displays in Experiment 4. These predictions were confirmed.

traditional interpretation, which has been formulated using the concept of figure–ground segmentation, and the interpretation provided by Boolean map theory.

The traditional interpretation is Bayesian in character: it states that figure–ground segmentation reflects the fact that it would be a great coincidence for two adjacent objects to happen to have outer contours that fit snugly into each other (e.g., Rock, 1983). Instead, it is overwhelmingly more likely that one of the curves only appears so because of occlusion (i.e., part of one object is actually hiding behind the other object). In recognition of this fact, the visual system determines which object is behind the other and declines to assign the object the shape dictated by the contour (in some formulations, it is represented as perceptually shapeless). This account attributes the failure to assign shape on one side not to any processing difficulty but rather to intelligent inference. We call it the *intelligent contour removal account.* Most work in the field has adopted this approach. One example is the parallel interactive model of configural analyses (PIMOCA; Peterson, de Gelder, Rapcsak, Gerhardstein, & Bachoud-Levi, 2000). This model assumes that different parts participate in perception organization in parallel but also tend to inhibit each other across edges. PIMOCA has been successful in explaining some pertinent phenomena (e.g., the role of familiar shapes).

Although the intelligent contour removal account has evident merit, its assumption about the direction of causality here may be questioned. We suggest that the human visual system cannot simultaneously process two abutting shapes because the spatial processing required for shape analysis relies on a Boolean map, and only as a consequence of this inability is a mechanism needed to decide which region is likely to be more important (figure). If the intelligent contour removal account is correct (at least in its strongest form), observers should be able to process the shapes on both sides when neither of the two regions appears to be behind the other. On the other hand, Boolean map theory predicts that even if neither of the two regions appears to be behind the other, the shapes on both sides can still not be accessed simultaneously.

As it happens, some features found in natural scenes do at least as good a job of illustrating the point as would an artificial display (while possibly also raising questions about the statistical assumptions underlying the intelligent removal interpretation). In the photographs seen in Figure 14, when inspecting the curves in the

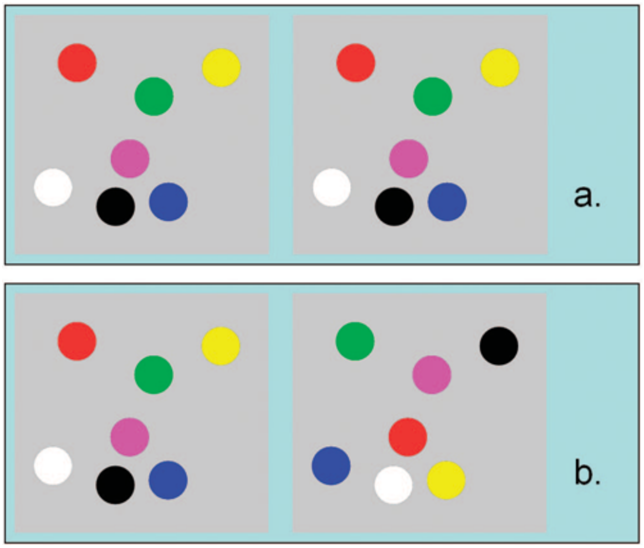


Figure 11. In Panel a, comparing the shapes (i.e., set of relative locations) of two patterns is at least partially parallel, whereas comparing the set of colors seems to require sequential checking from color to color, based on introspective report. Shuffling the colors around among locations, as in Panel b, does not seem to affect the ease with which the shapes can be matched. However, shuffling the locations around noticeably impairs the ability to verify that the set of colors appearing in the left display of Panel b is identical to the set of colors appearing on the right.

middle of the pictures (a ridge in Figure 14a and a crevice in Figure 14b), the reader will probably notice that other visual cues have been strong enough to overcome the tendency of figure–ground segmentation. That is to say, the visual system correctly determines that neither of the two sides is behind the other. However—at least judging from introspection—although it is easy to attend to both sides, extracting the shape of the center ridge or crevice can only be done from the left side or the right side, but never both.

This demonstration argues against the extreme version of the intelligent contour removal account. The account could doubtless be modified to explain this phenomenon. For example, even if the primary reason for assigning the shape information to only one region or the other is to deal with occlusion, this tendency may be misapplied to other situations in which occlusion is not present. In any case, the phenomenon is directly predicted by Boolean map theory and subsumed under a much more general principle.

Obligatory encoding of location. One basic tenet of Boolean map theory is that to access a feature, one must create a Boolean map encompassing a region comprising only that feature. Therefore, it is impossible to know the feature value of an object without also apprehending its location. This claim of Boolean map theory is similar to the claim emerging from research on detection of gratings at threshold, to the effect that signals are detected on *location-labeled channels* (A. B. Watson & Robson, 1981). The difference, however, is that according to the present view, the channel could be a potentially complex and rich spatial representation encompassing articulated and discontinuous regions.

Ordinary visual experience is surely consistent with this view. Whenever one sees a tree, a car, or a boy playing basketball, one simultaneously knows where it or he is with some reasonable precision. This is so intuitive that it may seem trivial. However, it is not at all self-evident and indeed, is not true of all sensory modalities; one can be perfectly aware of an auditory event such as a bird singing while having only a vague sense of its location (see Kubovy, 1981, 1988, and Kubovy & Van Valkenburg, 2001, for converging evidence that location is unique in vision, but not audition, and for stimulating discussion of this issue).

Indeed, feature-integration theory explicitly claims that features can be accessed without knowledge of their location (Treisman & Gelade, 1980). Although initial evidence appeared to favor this account, analyses that considered the role of guessing led to the opposite conclusion (Johnston & Pashler, 1990; for a recent review, see Quinlan, 2003; see also Driver & Vuilleumier, 2001). The literature is now generally consistent in implying that a feature cannot be accessed without simultaneous knowledge of its location.

No access to details within a Boolean map. Boolean map theory predicts that an observer will have no access to feature information that applies to only part of the Boolean map. When one selects the whole, one has no access to a feature description that pertains to only one part of this whole. For example, when one selects both objects in Figure 3, it is not possible to assign the color red to the whole Boolean map even if it is a single-feature value. This notion is consistent with the findings of He, Cavanagh, and Intriligator (1996). They found that even if the orientation of a crowded grating patch could not be explicitly reported, viewing the patch could nonetheless induce an orientation-specific adaptation effect, indicating that its orientation must be represented in the visual system. He et al. argued that this happens because attention cannot be allocated to the crowded individual grating.

Poverty of momentarily accessible information. One basic tenet of Boolean map theory is that observers can access only one single feature at a time. How can we reconcile that with the commonsense observation that people see so many things simultaneously? Perhaps people do not really have access to as much visual information as they are tempted to suppose.

In studies of change detection with intervening delays sufficient to produce flicker, it has been found that people detect only a small proportion of changes that are introduced into temporally successive displays separated by intervals of about 100 ms or longer (Levin & Simons, 1997; Pashler, 1988b; Rensink, 2000a, 2000b; Rensink, O'Regan, & Clark, 1997, 2000). As many researchers have pointed out, this appears to challenge the ordinary phenomenology of a rich visual awareness (although the term *change blindness* appears overdramatized, given that four to five items are often successfully processed in change detection designs; Pashler, 1988b). Regardless, change detection performance could either underestimate or overestimate the amount of information perceived at any given instant because the task may depend upon a memory that could encompass more or less than what is instantaneously perceivable.

One recent account of change detection (Rensink, 2000a, 2000b) proposed that observers access only one object at one time. By contrast, Boolean map theory claims that observers can access only one Boolean map, not one object, at a time. There is no widely accepted definition of one object. However, it seems to us that the stimuli in Experiment 3 (see Figure 10a,

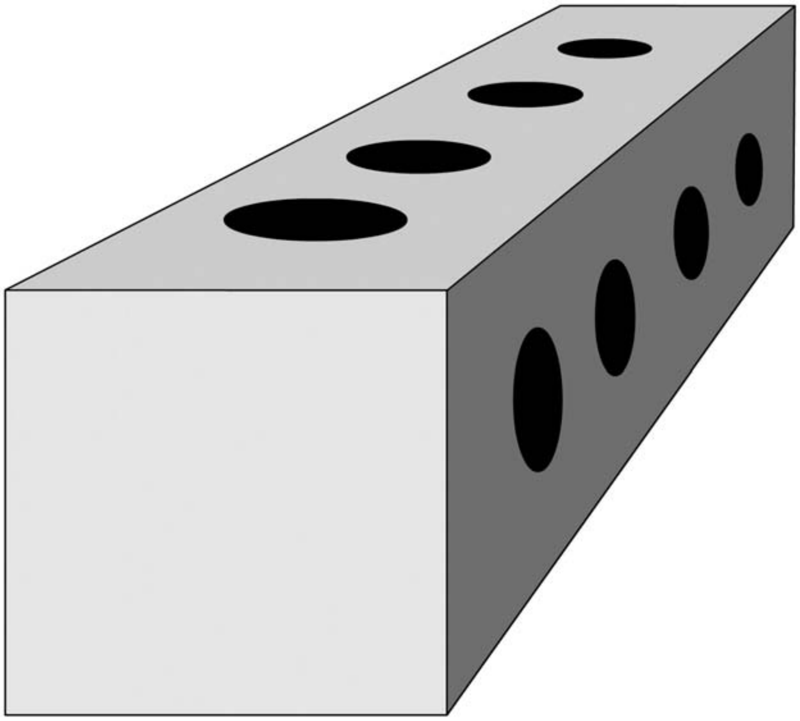


Figure 12. A demonstration that depth is an exception to the "one feature level at one time" principle of access, as illustrated in Figure 11. The black spots in various depth levels can be accessed simultaneously, and an observer can perceive a pattern in three-dimensional space.

a four-segment wheel) would usually be regarded as one object and the stimuli in Experiment 4 (see Figure 10b, a two-dot pattern) would usually be regarded as two objects. As shown above, different parts of the four-segment wheel have to be accessed sequentially, whereas the two-dot pattern can be accessed all at once. Therefore, it seems more plausible to suppose that it is the Boolean map, not the object, that corresponds to what can be accessed at one moment (see also Jiang, Chun, & Olson, 2004). Naturally, readers may or may not agree with our opinion on what constitutes an object, but in any case, the concept of Boolean map is at least a more clearly defined concept. In sum, change detection findings point up the remarkable poverty of the momentarily accessible visual information. However, views derived from change detection research about what can be accessed at one moment (an object) are open to question, whereas the view that what can be accessed at one moment is a labeled Boolean map seems more consistent with the (limited) existing data.

Feature Labels for Different Dimensions

Above, we discussed the proposition that there can only be one label per map for any one dimension (e.g., color). In the case of different feature dimensions (e.g., color and orientation), can there be a separate featural label for each dimension in a single labeled Boolean map? Concretely, this question amounts to whether observers can perceive features of different dimensions simultaneously (e.g., the redness and verticalness of an object). This question can be addressed by showing participants a few objects that can vary in two dimensions. If performance is worse when they have to report feature values in both dimensions, then that would indicate a difficulty in attaching two labels to the same Boolean map. The number of objects has to be very small and the perceptual task has to be very difficult (e.g., very brief displays) to ensure that what is measured is a perceptual limit instead of a limit of visual working memory. A few pertinent studies exist (e.g., D. A. Allport, 1971; Duncan, 1984; Magnussen & Greenlee, 1997; see also Egeth, 1966), and they agree in finding no competition between different dimensions. Thus, the answer to our question above appears to be in the affirmative: A Boolean map can have labels for different dimensions.

Given this specification, Boolean map theory predicts the now-classic finding of the same-object advantage (Duncan, 1984): Perceiving two features of one object is easier than perceiving two features that belong to different objects. The reason is because only one Boolean map has to be created in the first case, whereas two Boolean maps have to be created in the second case.

Boolean map theory claims that what is commonly called object-based attention is just another manifestation of the nature of Boolean maps. For two features to be simultaneously accessible, the crucial requirement is that they reside in the same region, not that they belong to one continuous whole. For example, Boolean map theory predicts that if two features belong to different regions of one object (e.g., Figure 10a), there will be competition between them even if they belong to the same object from the perspective of perceptual structure, as confirmed by Experiment 3.

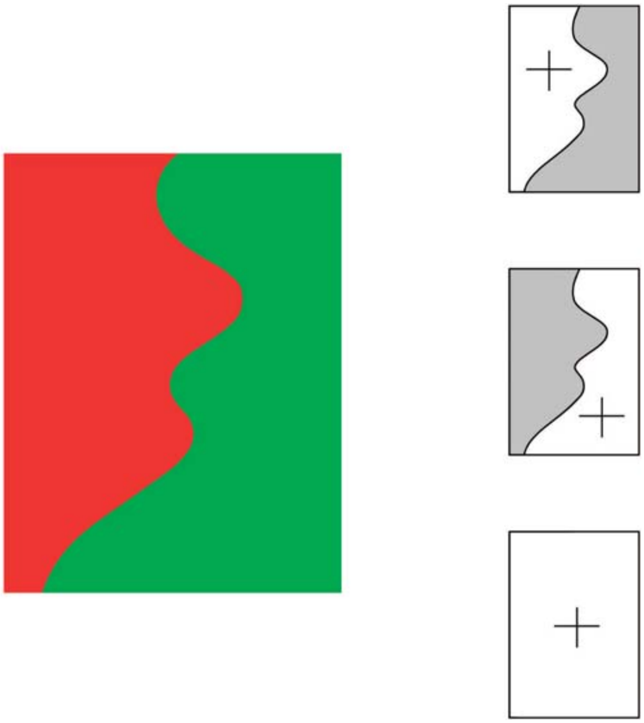


Figure 13. When a red and a green region abut each other, Boolean map theory predicts that selection of both regions entails a Boolean map that cannot support analysis of the shape of the contour that separates the two.

The Creation of a Boolean Map

In this section, we first clarify the two principles that have been proposed to characterize the creation of a Boolean map: the selection from one dimension and the application of Boolean operations. Then, we defend the tenet of feature-by-feature selection. Next, we defend the claim of the availability of Boolean operations, arguing that a selection relying upon more than one dimension is solved by a Boolean intersection operation and not by the use of a top-down salience map. Finally, we argue that all top-down control is mediated by creation of Boolean maps.

Selection From One Dimension: Selection Based on One Feature Value

Creating a Boolean map from one feature value within one dimension is the simplest and most fundamental way of creating a Boolean map.[10] There seems to be no question that this is possible when the feature space is clearly defined (e.g., selecting all red objects from among green ones), but we should briefly describe what can be counted as a feature dimension. As noted, we borrow the concept of feature maps from feature-integration theory (Treisman & Gelade, 1980; Treisman & Gormican, 1988; Treisman & Sato, 1990) and the guided search model (Wolfe, 1994). There is a certain amount of ambiguity in those theories as to what counts as a feature. It is not our purpose to shed new light on this issue; on the basis of prior work, a reasonably complete list might be color, orientation, size and spatial frequency, motion, depth, and perhaps certain aspects of shape (e.g., curvature, closure, digit/letter identity). Also, location itself can also be used for selection; for example, selecting a ball from an array of identical objects would require generating a Boolean map based on a specification of location.

In addition to the standard feature maps, we assume that there is also a bottom-up salience map. This map represents locations that are inherently salient based on a fixed set of computations, for example, those extracting local discrepancies between a feature and the surround (e.g., Itti, Koch, & Niebur, 1998). One such example, which we discussed earlier, is selecting all objects present in the display. Also, in Figure 15, it is fairly easy to see the pattern of all the nonblue colors because they are local minority. This cannot simply be based on a strategy of excluding one color: It seems extremely difficult to exclude one color in Figure 16. The concepts of salience and salience maps have often been used in the previous literature to refer to a map that combines information from various underlying feature maps in a way that reflects the individual's momentary goals and task set (e.g., Cave & Wolfe, 1990; Wolfe, 1994; Wolfe, Cave, & Franzel, 1989). For example, if a person searches for a red vertical object among green vertical objects and red horizontal objects, some theorists postulate a salience map that combines the redness and verticalness into a single scalar to assess the overall relevance of an object. This concept is termed *top-down salience map* in this article. This distinction is important because Boolean map theory denies the existence of a top-down salience map, and one goal of the present article is to persuade the reader that this reasonable-sounding notion is in fact highly questionable.

Can Observers Simultaneously Select Two Feature Values?

This question about possibility of simultaneous selection of two feature values is very fundamental. However, it has never been directly addressed. The most relevant evidence available from the visual search literature suggests that one can simultaneously select two features if they are linearly separable in color space from all other colors in the display (see Bauer et al., 1996, and D'Zmura, 1991, for evidence that this constraint operates in disjunctive search for a single color-based visual search target). However, this does not appear to govern performance in more general cases of attention-to-structure tasks. In Figure 16, it appears very difficult to select union (green, blue) or union (red, yellow), even though they are easily separated in color space from the remaining colors. So, it seems that one cannot divide the color space into two parts and select one of them; one really has to base his or her selection on one color at one time.

---

[10] One may point out that previous findings of automatic capture of attention when observers lack any explicit intention (Yantis & Jonides, 1984) are inconsistent with our notion that a Boolean map is always created from one feature value. However, there may well be one or more default strategies for creating Boolean maps (Folk, Remington, & Johnston, 1992, 1993; see also Pashler & Harris, 2001; Yantis, 1993): for example, creating a Boolean map from the bottom-up salience dimension (favoring the most salient object; Yantis & Egeth, 1999; Yantis & Johnson, 1990). Also, there is some reason to doubt that attention capture is really automatic (Koshino, Warner, & Juola, 1992; Warner, Juola, & Koshino, 1990).

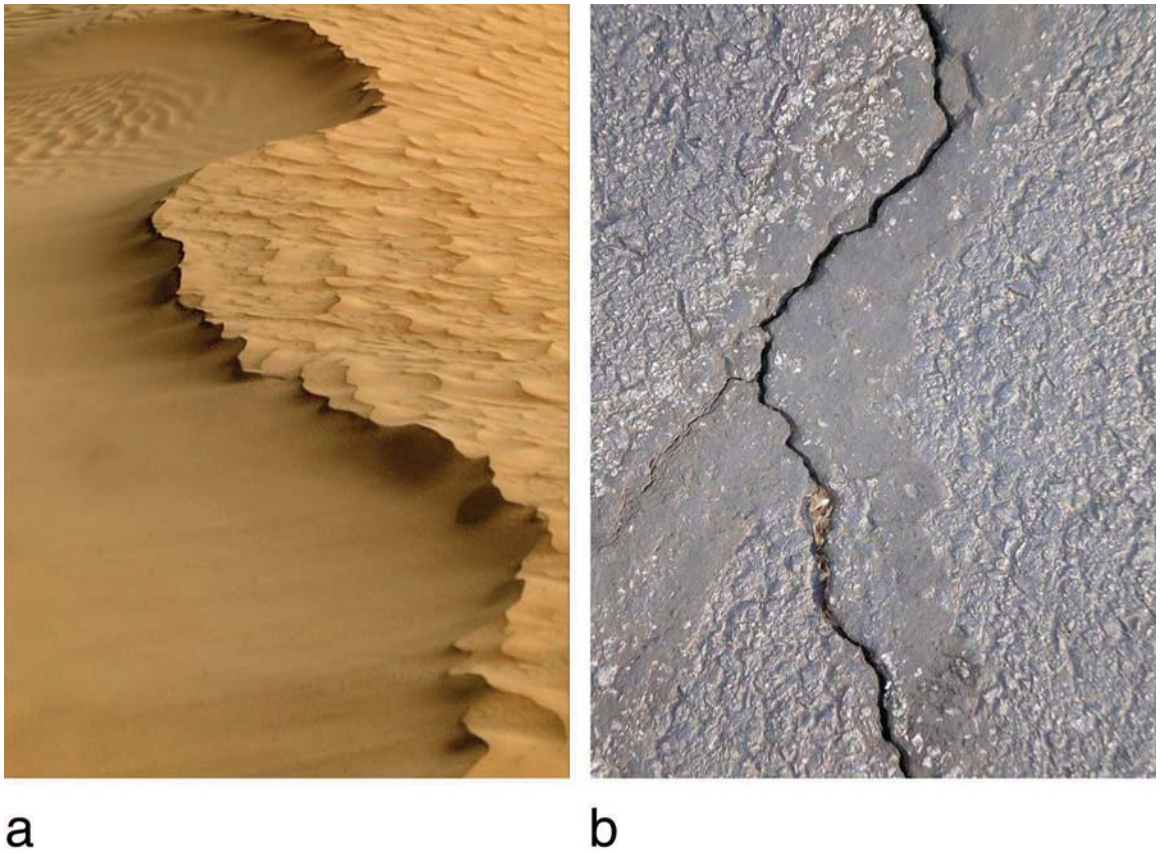


Figure 14. Shape extraction and figure–ground segmentation. Even when relative depth cues are absent, as in these photographs, the difficulty in assigning the central contour (a ridge in Panel a and a crevice in Panel b) to determine the shape of both surrounding regions appears to persist, as Boolean map theory would predict. Photo in Figure 14a copyright 2007 by Declan McCullagh/mccullagh.org. Used with permission.

As mentioned earlier, selection by location differs from other forms of feature-by-feature selection. For one thing, location-based selection is not limited to a small spot but can be also effectively applied to a region (i.e., the objects in that region) even when these are more or less widely distributed in space. The exact principles governing this location-based selection should be studied in the future, but for present purposes, we need only assume that slightly more flexibility exists in this process.

The preceding discussion points up the fact that questions about attention to structure are not trivial corollaries of the sorts of questions posed in visual search studies. The phenomenon also has potential practical implications for practical fields like information visualization, where the question of what can be seen at a glance (in cases much like Figure 16) is a critical factor in designing new technologies for data exploration and presentation (Bertin, 1983; Ware, 2000). It should also be noted that the knowledge gleaned from the literature on texture segregation is insufficient to characterize the limitations governing perception of structure. A standard texture-segregation task (Julesz, 1981) involves determining when one (usually convex and regular) region can be perceptually separated from the rest of the pattern. In Figure 17, texture segregation of blue and green from red and yellow is apparently fast and effortless, in clear contrast to the general difficulty revealed in attention to structure with the same color choices. For data visualization, the situation of Figure 16 is clearly more relevant than the situation in Figure 17. Thus, examining tasks such as those in Figure 16, along with more standard texture-segregation tasks, should help us begin to achieve a more complete understanding of attention to structure.

Boolean Operations as the Only Way of Combining Different Dimensions

The purpose of this section is to argue that the strategy described by the second principle, combining the output of one-dimension selection with the present Boolean map in Boolean operations, is the only way that selection can be governed by information from more than one dimension. To put it another way, a Boolean map can be created only by giving one control value at a time to the *FL* (sometimes applied iteratively, so as to modify the current Boolean map).

The Boolean operations: Intersection or union. Boolean map theory contends that people are able to apply certain Boolean operations to modify existing Boolean maps. Using the operation of union means creating a new Boolean map consisting of everything that either (a) is presently attended or (b) satisfies some newly specified property. Using the operation of intersection means creating a new Boolean map consisting of everything that

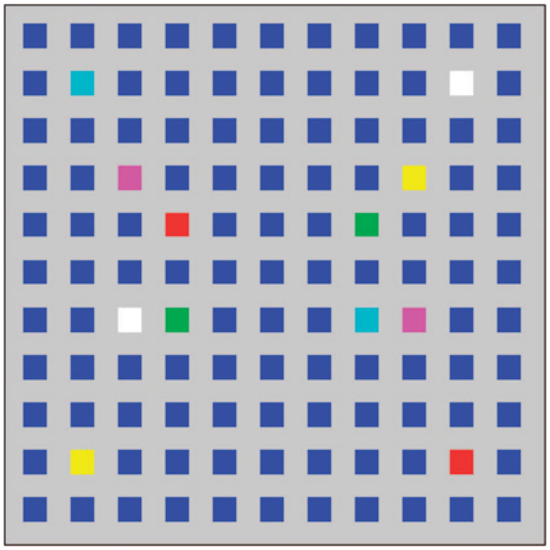


Figure 15. A symmetric structure is readily detected when only a few elements differ from a homogeneous surround. This is interpreted as evidence that a bottom-up salience computation based on local differences can specify a Boolean map without any single unifying feature.

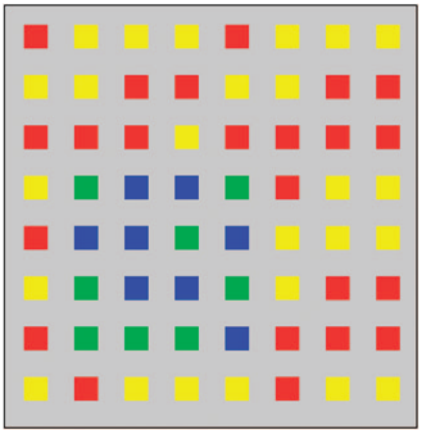


Figure 17. The compact blue-or-green square is readily segregated from the red-and-yellow surround in the classic perceptual task used to study texture segregation. The contrast between this successful texture segregation and the poor perception of structure seen in Figure 16 is argued to demonstrate that perception of structure, as discussed in the present article, should not be equated with texture segregation.

both (a) is presently attended and (b) satisfies some newly specified property. It should be mentioned that postulating these Boolean operations does not contradict the claim that at any given instant, there is only one Boolean map underlying momentary visual awareness because the new Boolean map is assumed automatically to replace the old one.

This potentiality to use Boolean operations to govern conscious access to structured displays can be readily demonstrated, although interestingly, it has not to our knowledge been remarked upon previously. In Figure 18a, the reader is invited to start by seeing all the red items and then to select from these just the circles. We contend that this is achieved by first creating a Boolean map consisting of all the red items and, from there, utilizing intersection, as described above, to create a new Boolean map consisting of just the red circles. From this new Boolean map, one can effectively extract spatial information: For example, an observer sees that the red circles are arranged in a symmetric pattern, whereas, for example, the red crosses or squares are not. As we argue below, the task of conjunction search (Treisman & Gelade, 1980) is also accomplished by intersection—only, in this case, just one item is left after the application of the second step.

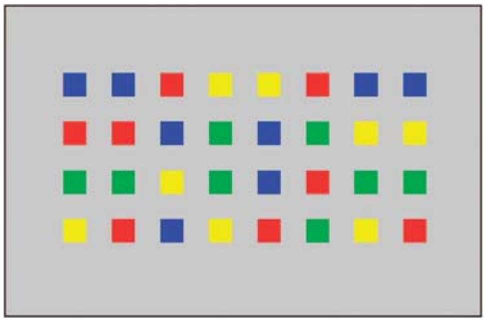


Figure 16. Selecting two colors simultaneously. Selecting the union of the green and the blue or the union of the red and the yellow is very difficult—despite the fact that the first pair of colors is linearly separable from the second pair in CIE color space.

In Figure 18b, most observers report that they can (with some effort) select the union of the red pattern on the left and the coarse texture pattern on the right. Alternatively, one can select the union of the red pattern on the left and the fine texture pattern on the right. In each case, the spatial arrangement of the union can be vividly perceived. Tasks such as selecting a few from many identical objects seem to be also accomplished by iterative union operations. The union operation seems to have one striking restriction. As shown in Figure 16, a union operation cannot be used to select union (red, yellow) together. We speculate that the union operation is effective only when the two sets to be combined are not spatially intertwined. This is perhaps because the union operation is implemented by connecting two global shape descriptions into one. Naturally, two spatially intertwined patterns will make a completely new shape instead of a simple connection of the two individual shapes, so the union operation is foiled.

Certainly, we do not intend to suggest that the visual system can create Boolean maps of unlimited complexity through repetitive Boolean operations of intersection and union. It appears that there is some limit as to how long and elaborate a sequence is possible. Plausibly, this is because the sequence of Boolean operations used to create one Boolean map must be stored in some machinery (which we might clumsily refer to as the *Boolean map creating procedure,* or BMCP). For example, green → intersection vertical is stored in the BMCP if one selects vertical objects among the green objects, or $(x_1, y_1) \rightarrow$ union $(x_2, y_2) \rightarrow$ union (x_3, y_3) is stored if one selects three objects sequentially by location. Plausibly, this BMCP is a form of working memory, with very limited capacity, therefore allowing only a few intersection and union operations to be composed in series (e.g., pulling together arbitrary objects

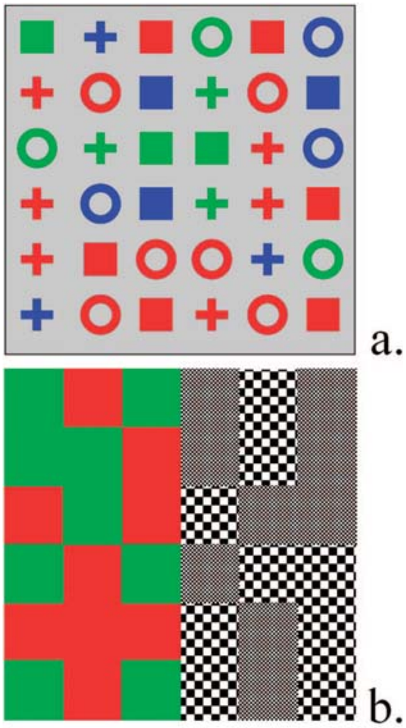


Figure 18. a: Boolean intersection task—first, select all the red items, then see the subset that are Os. b: Boolean union task—first, select the red, then additionally see the coarse texture on the right. Unlike Figure 15, the nonoverlapping union supports Boolean mapping (albeit with some modest but apparent difficulty).

through repetitive union operations can evidently only encompass 4–5 of them).

Boolean operations and top-down salience map: Two competing accounts. Boolean map theory claims that if a judgment requires combining information from more than one dimension, a Boolean map can be created from only one single-feature dimension at a time, and selection based on a combination of two dimensions is accomplished by the intersection operation mentioned above. This entails a subset search strategy for the classic color–form conjunction search task of Treisman and Gelade (1980). For example, if one wants to find a red vertical target, a Boolean map must be created to mark all the red objects. Then, this Boolean map can be combined with information from an orientation map via the intersection operation to generate a new Boolean map that contains the vertical object among them.

Various previous theories (notably, Wolfe, 1994) have assumed that information from different feature dimensions is combined into a master salience map. This salience map is a combination of a bottom-up salience map and a top-down salience map. As mentioned above, the present formulation agrees that the bottom-up salience map plays a role in selection. The crucial difference between bottom-up salience and top-down salience is that the former is not task specific and is not under voluntary control. Thus, our claim can also be understood in this way: There is a salience map in visual processing, but it is strictly bottom-up and cannot be altered by top-down control, for example, by instructions or voluntary goals.

The crucial difference between intersection (subset search) and top-down salience is that in the intersection operation, the feature map is combined with the preexisting Boolean map, not directly with another feature map. One might think the difference is trivial. However, as we show below, even if these two models make relatively similar predictions about standard conjunction search, they make starkly different predictions in other situations.

In fact, there has for some years been empirical evidence in the literature pointing to the existence of a subset search strategy in conjunction search (Egeth, Virzi, & Garbart, 1984; see also Kaptein, Theeuwes, & van der Heijden, 1995). Egeth and colleagues (1984) showed that in this task, search is mainly restricted to target-color-bearing distractors and that search latencies are hardly affected by the number of target-orientation-bearing distractors. However, their study had the weakness of not controlling the relative strength of the feature difference in the two dimensions. Therefore, a master top-down salience map could also potentially explain the findings of Egeth and colleagues by maintaining that the top-down salience is much greater for a target-color distractor than for a target-orientation distractor, thus attention is always guided there.

Subset search is a necessary assumption, whereas top-down salience is not. Additional support for the subset search strategy is seen with displays such as those in Figure 19. It is easy to find the orientation singleton among just the green bars. This must be done by selecting green items first. If features are combined together to form a top-down salience map, then orientation signals

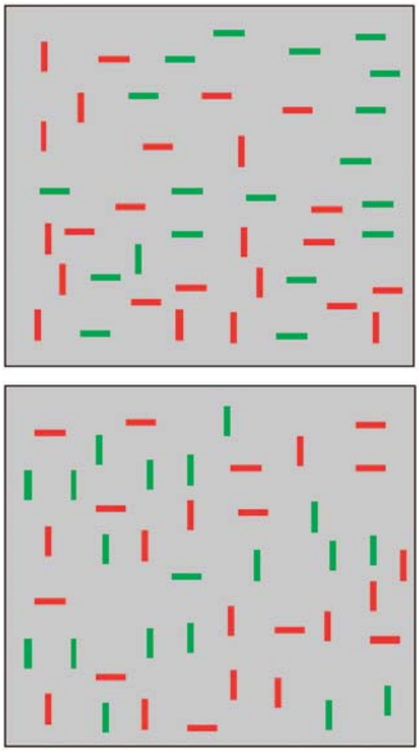


Figure 19. Subset strategy in conjunction search. Task: Find the orientation singleton among just the green bars (the orientation of this singleton being unspecified). Boolean map theory predicts this can readily be achieved, whereas some other theories appear to predict the opposite.

will not be useful because it is not known in advance which orientation should be boosted. Wolfe (1994; see also Friedman-Hill & Wolfe, 1995) admitted the necessity of a subset search in such displays; he argued that this was one option, with another option being combining feature information into a top-down salience map. However, the evidence that has been adduced in support of this latter strategy seems open to question.

One example of such putative evidence is the finding of better performance in triple-conjunction search (e.g., finding a target that is red and horizontal and large) as compared with double-conjunction search (Wolfe, Cave, & Franzel, 1989). This advantage is also naturally explained by Boolean map theory. In a typical triple-conjunction search task, creating a Boolean map from any one feature dimension excludes two thirds of the distractors. In typical double-conjunction search task, creating a Boolean map from one feature dimension filters out only half of the distractors. Thus, there are 50% more relevant items in double-conjunction search than in triple-conjunction search. In addition, the participant only has to exploit any two of the three relevant feature dimensions. This wider range of strategy choices could also aid performance. Moreover, the local featural difference between items is larger in the case of triple-conjunction search, and thus, bottom-up salience could contribute to the performance. Wolfe (1994) also argued for the master salience map on the basis of a comparison of the slopes for target-present versus target-absent trials. However, if one assumes that observers apply the subset search strategy serially for different parts of the display, these patterns can be explained.

Taken together, the data show that subset search is a necessary notion, whereas the top-down salience map is an unnecessary notion. This does not directly disprove the possibility of top-down salience computation, of course, but does undermine the idea on grounds of parsimony. Below, we present evidence that more directly challenges the existence of a top-down salience map.

Trading one dimension against another. If signals from various dimensions can be combined into the one single top-down salience map, it ought to be possible to arrange for one dimension to trade off against another in a visual selection task. Concretely, this means that observers should be able to simultaneously pull out objects that have a specific color or a specific shape by assigning positive salience weights to that color and that shape. Is this in fact possible? When looking at Figure 20, observers seem to agree it is extremely difficult to perceive the pattern of union (red circles, green crosses)—that is, disjunction. On the other hand, this difficulty does not happen simply because the red–green difference and circle–cross difference are not large enough as it is fairly easy to perceive the pattern of (green circles)—that is, conjunction. If top-down determination of salience is the underlying mechanism used to accomplish such selections, the salience difference between these two subsets ought to be the same regardless of which one is to be selected: In one case, green and circle would be boosted; in the other case, red and cross would be boosted. Therefore, the idea of a top-down salience map cannot account for the marked difference in the difficulty of selecting these two subsets. Boolean map theory, on the other hand, offers an explanation of this difference: The selection of (green circles) is carried out through intersection, whereas the selection of union (red circles, green crosses) is difficult because the union operation is not effective when the two sets intertwine. One may note, of course, that the selected set is homogeneous in the case of conjunction but heterogeneous in the case of disjunction and argue that this explains why the task is more difficult in the case of disjunction than conjunction. However, the very fact that homogeneity or heterogeneity of the actual feature values (and not just the description that governs the selection) makes such a big difference proves that the selection is implemented in a way that depends upon the actual feature values present in the display. There is no reason this should be the case according to the top-down salience map hypothesis. Even if the original feature values are heterogeneous in the case of disjunction, the top-down salience map should have combined them into one single scalar, and thus, the success of selection should be comparable for disjunction and conjunction.

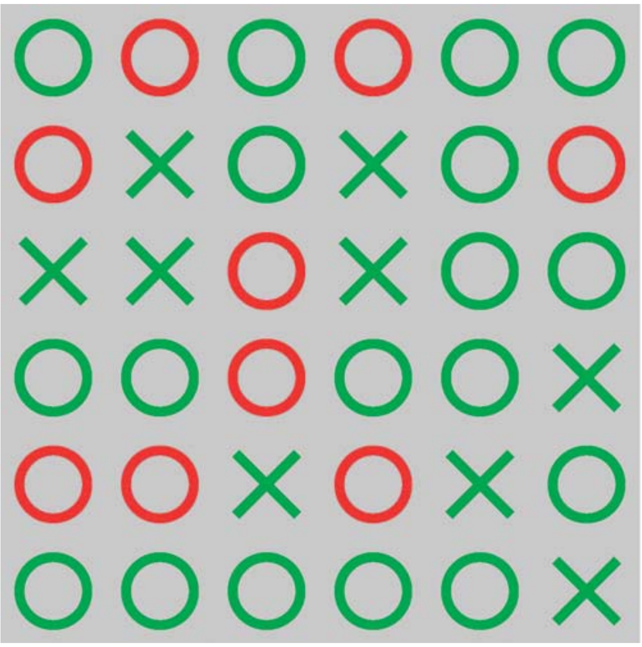


Figure 20. It is easy to perceive the pattern of green circles, whereas it is extremely difficult to perceive the union of the red circles and green crosses. A top-down salience map, if it exists, should allow one to separate these two subsets equally well regardless of which one is selected. This challenges the view that the binding problem is solved by a master top-down salience map.

Similarly, Figure 21 illustrates an example in which the targets are defined as elements with a high combined size–luminance value (compared with the distractors). Both size and luminance vary in four levels, so the targets do not have any particular predetermined size or luminance values that can be searched for. However, if size and luminance signals could be added to generate one single top-down salience measure, the task should be relatively easy. By way of analogy, suppose one is reviewing files of graduate school applicants and wishes to prescreen by two factors: GRE scores and grade point average (GPA). The most convenient way to do this task is to compute a composite index (analogous to top-down salience) and rank them accordingly. If one is not able to do that, one may have to look for particular combinations (e.g., GPA > 3.0 & GRE > 1400; GPA > 3.3 & GRE > 1300; GPA > 2.7 & GRE > 1500), which is likely to be a very laborious process. Going back to the visual search task of Figure 21, if the top-down

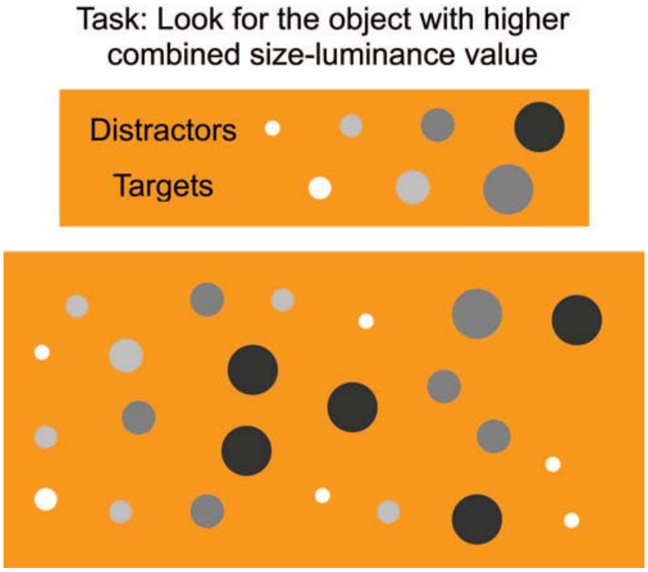


Figure 21. The top panel defines targets and distractors. The reader is invited to search for these targets in the larger bottom panel. If a flexible top-down salience map can be constructed on the fly, then it should be possible to find targets defined as elements with high combined size–luminance values relative to distractors.

salience map allows one to search for a high combined size–luminance value, then the task should be relatively easy. If no such mechanism exists, however, one is forced to search for the particular pairs of size and luminance one after another, and the task should be very laborious. In the bottom panel of Figure 21, there are three targets; our informal observations suggest that observers do indeed find the task very difficult. We conclude, therefore, that when observers perform a conjunction search, they really have to search for a particular conjunction of feature values, not an abstract size–luminance value without concern for the concrete values of size and luminance.

All these phenomena make good sense if there simply is no top-down salience map that can be used to allow two dimensions to trade off against each other. Of course, one might propose additional restrictions on the top-down salience map to account for these results. For example, one might assume that the top-down salience map works only in the case of conjunction and not in the case of disjunction. This would make the notion of top-down salience map very similar (for the purposes of this task) to the intersection operation in Boolean map theory (see also Treisman & Sato, 1990, for a revised version of feature-integration theory that has something in common with this view). The only remaining crucial difference is that such an account would propose that red–vertical targets are directly pulled out with no intermediate step, whereas Boolean map theory predicts that one has to first select everything red or everything vertical. We try to distinguish these two possibilities below.

Special difficulty for conjunction in brief displays. If a top-down salience map allows a conjunction to be directly selected much as is a feature, albeit less effectively, then the time course of conjunction search should be very similar to an inefficient feature search of similar difficulty. However, Boolean map theory predicts that the time course of conjunction search should have a distinctive property. The reason for this is that in conjunction search, the chief difficulty of the task lies in the fact that two Boolean maps have to be created successively. Therefore, the time course of conjunction search in a brief exposure should have a spoon shape, with performance increasing extremely slowly at the very beginning but then increasing sharply after a certain point. The reason is as follows: Conjunction search is a two-step process, in which the first step (creating the first Boolean map) provides almost no useful guidance for selecting a response in the task. On the other hand, even a difficult feature search is a one-step process in which useful information might be expected to emerge from the very outset (albeit slowly). By way of an analogy, a young person graduating from high school often has the option of going to college for 4 years and then finding a job. The 4 college years do not directly provide any income (analogous to the first step in conjunction search). The individual can also go directly to work, typically providing an immediate but smaller income (analogous to the difficult feature search). Thus, if one measures the accumulated income versus number of years after high school, the direct-to-work strategy (i.e., difficult feature search) yields greater benefits for the first years, but eventually the college strategy (i.e., conjunction search) will outstrip it. Naturally, the turning point varies from trial to trial, smoothing out any sharp elbows in the average data. Nevertheless, if one compares the time courses of conjunction search and an inefficient feature search of appropriate difficulty, one would expect a crossover interaction, with the performance of the conjunction search being worse at the very beginning but ultimately exceeding performance of an inefficient feature search.[11] Experiment 5 confirms that the predicted crossover interaction can in fact be observed (see Figure 22). A top-down salience map—or any other mechanism that assumes the selection of a conjunction is functionally similar to but only less efficient than the selection of a feature—would not seem to predict this pattern.

Recent studies on feature-based attention (Moore & Egeth, 1998; Shih & Sperling, 1996) also indicate that signals from different feature dimensions cannot be combined in very brief displays.[12] For example, in Moore and Egeth (1998), participants searched for a digit in a display composed of equal numbers of red and green letters. In the constant condition, the color of the target was constant and known to participants. In the random condition, the color of the target was randomly specified on each trial. Not surprisingly, performance in the constant condition was better. If signals from different feature dimensions are combined into a top-down salience map, then this advantage presumably comes from giving some weight to the informative color dimension when

---

[11] In the experiment described here, to match the overall performance, the feature search was deliberately chosen to involve much lower discriminability than the conjunction search, so the time course could be compared meaningfully. The classic finding of steeper slopes for conjunction search pertains to experiments in which the same feature values are used in both tasks.

[12] Moore and Egeth (1998) did not interpret their result in this way. Their main purpose was to show that color-based attention does not facilitate the sensory processing of the assigned color directly but gives priority in visual search to the items of the assigned color.

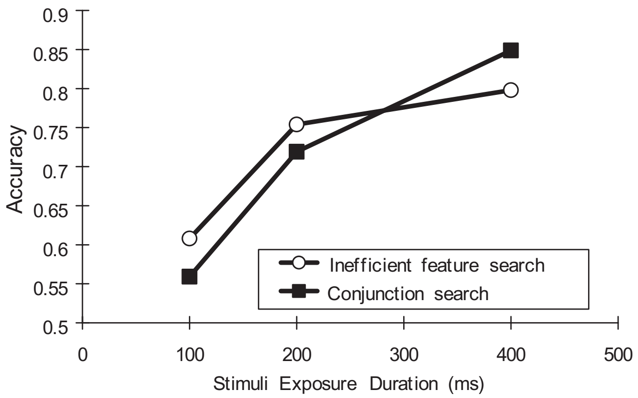


Figure 22. Proportion correct as a function of exposure duration of a search display (followed by mask) in Experiment 5. This experiment compared a color–form conjunction search and a relatively difficult feature search. Boolean map theory entails that two Boolean maps must be created in sequence to perform conjunction search, whereas in feature search, this is not necessary. Therefore, the conjunction search should take a much longer time to initiate. Thus, the theory predicts that a crossover interaction may occur for appropriate difficulty levels, as the results confirmed.

computing the overall top-down salience. For example, suppose a red letter has a value of 10, the red digit has a value of 20, and a green letter has a value of 0, each with some noise. This would obviously help the red digit win in the competition.

If so, one would naturally expect the advantage of knowledge about target color to occur even with very brief displays. However, Moore and Egeth (1998) found that performance in the constant condition and in the random condition was identical for very brief displays. One may say that the signal reflecting the color difference was too weak in very brief displays to guide attention effectively. However, the signal from the digit–letter identity difference was strong enough to produce reasonable performance (about 0.75). Thus it is implausible that color difference, which was obviously a much stronger signal in their experiment than the signal from digit–letter identity difference, would be too weak to make any difference at all. On the other hand, Boolean map theory explains the results naturally: There was not enough time with the brief displays to accomplish the two-step sequence described. Selection on the color dimension was simply abandoned because it alone was not sufficient to locate the target.

Dissociation between top-down salience and bottom-up salience. Various published data on the distinction between what have been termed *singleton detection mode* and *feature detection mode* (e.g., Bacon & Egeth, 1994) also raise problems for the concept of a top-down salience map. In most proposals for a master salience map, top-down salience and bottom-up salience are assumed jointly to determine the tendency of an object to attract attention. However, previous studies on the distinction between feature search mode and singleton search mode suggested that these two types of signals are processed in a fundamentally different way. Pashler (1988a; see also Theeuwes, 1992) showed that search for a singleton is disrupted by the presence of another irrelevant singleton. Bacon and Egeth (1994) convincingly demonstrated that this happens only when the visual search is executed in a singleton detection mode (looking for something that is different from the rest); when search is performed in feature detection mode (e.g., looking for red), the irrelevant singleton does not significantly disrupt performance. We interpret this as indicating that when participants are in singleton detection mode, the Boolean map is created from the bottom-up salience map, so an irrelevant singleton will disrupt search. However, when an observer is in feature detection mode, the Boolean map is created from the relevant feature map, and thus, the irrelevant singleton will not disrupt performance. If top-down salience and bottom-up salience are always combined into one map, then even feature search should always be disrupted by the presence of another irrelevant singleton. Naturally, one can assume that the top-down salience map and the bottom-up salience map are functionally dissociated from each other. Such a modification, however, throws further doubt on the plausibility of top-down salience.

All Top-Down Control Is Ultimately Attributable to Selection

We have conceptualized selection as involving the creation of a Boolean map, but that is still one step from warranting the claim that all top-down control is ultimately attributed to creation of a Boolean map: After all, there could be top-down control other than selection. We argue that such top-down controls are all ultimately mediated by creation of Boolean maps.[13]

One important case to consider is change of perceptual structure (i.e., imposing a different percept on the same stimuli). For example, switching from the vase to two faces is one well-known example. Boolean map theory predicts that one can only change perceptual structure by creating an appropriate Boolean map (i.e., imposing one or another perceptual structure is implemented by allocating spatial attention in one way or another, not through any separate mechanism; see Slotnick & Yantis, 2005; Tsal & Kolbet, 1985), but it does not necessarily dictate what available strategy will be used in any given instance. A natural idea would be that creating a Boolean map encompassing an element that is more meaningful in the desired perceptual structure but less meaningful in the present perceptual structure tends to force a switch to the desired structure.[14]

In the case of face–vase demonstration, the pattern can be perceived as one vase or two faces placed on a background. The important fact is that if the display is now organized as two faces on a background, then the center region (vase-shaped) is not one element of such organization. Attending to that region forces the perceptual structure to switch the other way in which the center red region is indeed an element. Therefore, selecting the vase-shaped region tends to make that region be figure, and vice versa. This is supported by empirical evidence (e.g., Vecera, Flevaris, & Filapek, 2004).

The same point seems to be true in the case of other reversible pictures. In the classic young woman–old woman reversible fig-

---

[13] There are other studies viewed as instances of top-down control. It seems to us these studies can be viewed as involving either switching between different strategies for the creation of a Boolean map (e.g., attentional set effect for spatial frequency: Davis & Graham, 1981; Davis, Kramer, & Graham, 1983) or selection in the competition for postperceptual processes (Maruff, Danckert, Camplin, & Currie, 1999; Remington & Folk, 2001). Therefore, they do not contradict our claim.

[14] By using the word *meaningful,* we suggest that one part is functionally independent and functionally important. This ultimately implies a likelihood measurement from a Bayesian perspective.

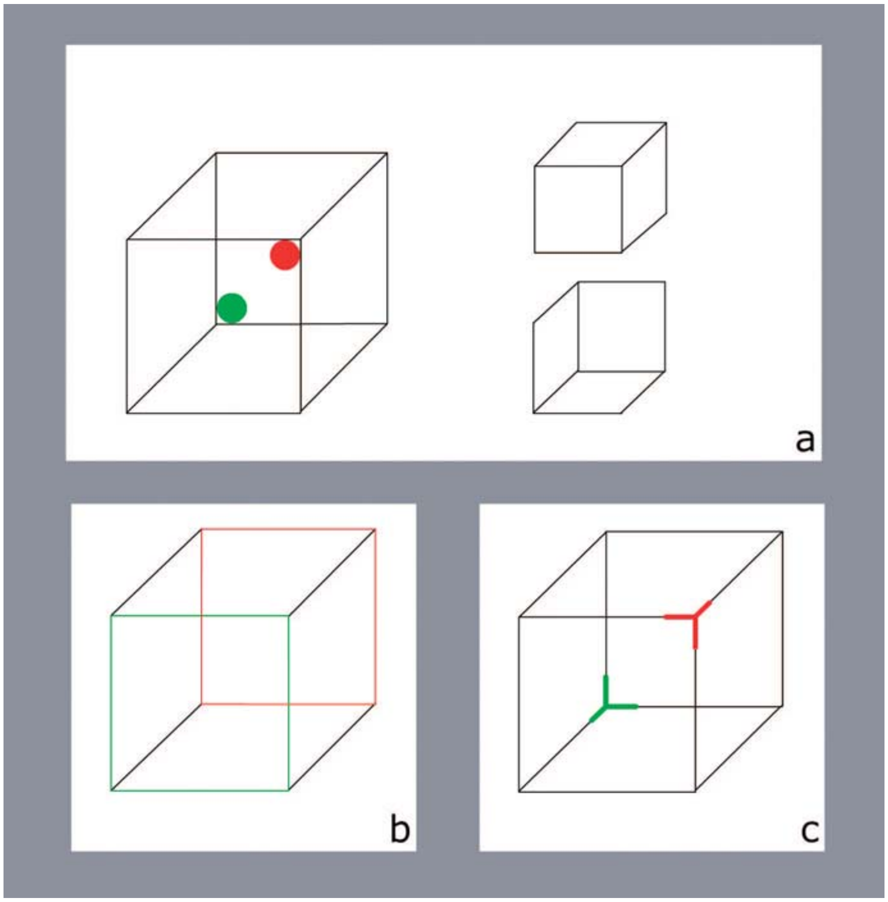


Figure 23. In a Necker cube, attending to one location or another (red vs. green dot) tends to force the cube to be perceived in one way or another (Panel a). This probably reflects the effect of creating a Boolean map to select part of the cube (e.g., surfaces, edges, or corners) instead of a spotlight covering the general region. Attending to the green surface (Panel b) seems often to have the opposite effect from attending to the green corner (Panel c).

ure, the display is organized differently when it is viewed as old or young. If one wants to switch, for example, from the young woman to the old woman, one attends to the region that would be the face of the old women but is the face and neck of the young women (face evidently being a more meaningful part than face and neck).

Similarly, with a Necker cube, attending to one location or another tends to force the cube to be perceived as one way or another (Kawabata, Yamagami, & Noaki, 1978; see Figure 23). This probably reflects the effect of creating a Boolean map to select part of the cube (e.g., surfaces, edges, or corners). For example, attending to the green surface in Figure 23b seems to have the opposite effect from attending to the green corner in Figure 23c. Therefore, the effect cannot be adequately explained by shifting fixation to one area or another because attending to both a green surface and a green corner would then involve shifting to the bottom left part of the cube. Selecting part of the cube (e.g., edges, corners, surfaces) with spatial attention (i.e., Boolean map) must be the crucial factor here. It would appear that the underlying rule may be that whatever part is selected by a Boolean map tends to be perceived as nearer to the observer (front side). One may surmise that an element in the front side is probably more meaningful than an element in the back side. Therefore it is also consistent with the notion that creating a Boolean map that is a more meaningful element of desired perceptual structure but is less meaningful in the present perceptual structure tends to force the perceptual structure to switch to the desired structure.

This account might make reasonable ecological sense. If the pre-Boolean map vision achieved two equally likely perceptual structures and there was doubt which one was the correct interpretation of the external world, top-down control might select one of these two interpretations. If a Boolean map is created for one part that is more meaningful in Structure 1 but less meaningful in Structure 2, then the visual system might reasonably take that as a vote for Structure 1.[15]

In sum, in various reversible figures, imposing a new structure seems to depend on moving spatial attention (i.e., creating a Boolean map) to the pertinent part of the object, supporting our claim that top-down control is achieved by creation of a Boolean map.

---

[15] This top-down influence probably has a noticeable effect only in the case of ambiguous perceptual structures, and this analysis should not be taken to imply that top-down control is either sufficient (i.e., that it can always decide the perceptual structure) or necessary (i.e., that reversals of perceptual structure do not occur spontaneously).

Relation to Previous Theories

At the outset of this article, we remarked upon some differences between Boolean map theory and prior theories. The next section compares Boolean map theory with several previously proposed theories, briefly highlighting some of the similarities and differences. From the perspective of Boolean map theory, theories of visual attention can be roughly divided into two categories: theories illuminating selection and theories dealing with conscious access.

Theories Illuminating Selection

Boolean map theory obviously owes a great debt to feature-integration theory (Treisman & Gelade, 1980) chiefly—but not only—with respect to Treisman's fundamental insight that the visual system has a severe problem in maintaining multiple associations between different spatial and featural representations. This, of course, is something that is often called the *binding problem* in the broadest sense. We have already described the two fundamental disagreements between Boolean map theory and feature-integration theory: First, Boolean map theory holds that only one feature value can be accessed at one instant, whereas feature-integration theory holds that multiple feature values are simultaneously available. This contradiction is due to a lack of explicit distinction between selection and access. The experiments presented by feature-integration theory prove only that feature values are simultaneously available for the mechanism of selection, in the sense that all of them are ready to be called upon. They are, however, as we have shown, not simultaneously available for conscious access in the sense that even if all of the feature values are ready to be called upon, only one can be actually pulled out. Second, Boolean map theory holds that a feature value is always accessed together with its location value(s), whereas feature-integration theory holds that a feature value can be accessed without its location value. On both issues, empirical evidence favoring Boolean map theory has been discussed. On the other hand, Boolean map theory also readily explains the original findings that feature-integration theory was constructed to explain. For example, in the case of conjunction–feature search distinction, Boolean map theory predicts that conjunction search should generally be more difficult than feature search because an extra intersection operation is necessary.

The guided search model (Wolfe, 1994) is a major alternative to feature-integration theory proposed to address a similar set of questions. The guided search model incorporates a number of ideas that have both interesting resemblances to and points of contrast with Boolean map theory. Like Boolean map theory, guided search views selection as inherently location based. On the other hand, the guided search model shares the same ambiguity about access as feature-integration theory and therefore says little about tasks other than visual search. As for the strategy of exploiting different dimensions in selection, guided search claims that multiple separate feature dimensions can be summed into one scalar array (salience map), which in turn determines the allocation of visual attention. Boolean map theory denies that a simultaneous weighting of different dimensions is possible.

Boolean map theory also makes a strong claim that selection can be on only one feature value at one time. This question is not formally discussed by the major previous theories illuminating selection and appears to be a novel feature of the Boolean map theory.

Theories About Access

The main theories that explicitly discuss access are the theory of target–target competition (Duncan, 1980a, 1980b) and the biased competition model (Desimone & Duncan, 1995). Duncan (1980a, 1980b) addressed access very explicitly and clearly. Boolean map theory follows Duncan in postulating a highly limited capacity of visual conscious access. The major advance to which Boolean map theory aspires is in providing a formal specification of constraints on access. Duncan (1980a, 1980b) stated that the access is subject to severe capacity limitations, whereas Boolean map theory seeks to characterize these limitations in the language of data structures.

The proposal of object-based attention (Duncan, 1984) highlights another side of visual attention, which feature-integration theory and other theories have failed to address. The major finding in Duncan's studies is that feature values of different dimensions from the same object can be simultaneously accessed, showing little competition with each other. Moreover, studies of object-based attention have shown that attention cannot be viewed as a convex spotlight that moves from one location to another. Both conclusions follow directly from Boolean map theory. The major advance to be hoped from Boolean map theory here is that the theory offers a more explicit notion of what it means for features to be *within object*—a notion that differs from conventional interpretations of objecthood (Scholl, 2001; Scholl & Pylyshyn, 1999; Scholl, Pylyshyn, & Feldman, 2001). When two features belong to different parts of a continuous perceptual whole, some may consider them to be within the same object. Boolean map theory instead claims two features can be accessed without competition only when they reside in the same spatial region. As shown earlier in this article (Experiments 3 and 4), this prediction of Boolean map theory receives empirical support.

The idea that observers can extract only surprisingly impoverished information from a glance at a scene is in some sense quite old: It is implied by the most widely discussed theories of visual attention (Broadbent, 1958; Duncan, 1980a, 1980b; Treisman & Gelade, 1980). However, the idea has been reinforced by the observation that people do a very poor job of detecting changes introduced over offsets of about 100 ms or longer whether the displays consist of arrays of familiar characters (Pashler, 1988b) or pictures of natural scenes (e.g., Rensink, 2000a, 2000b). A number of researchers, notably Rensink et al. (1997, 2000), have argued on this basis for the proposition that the conscious awareness of scenes is impoverished in content. Boolean map theory echoes this observation but places it within a perspective that also makes sense of the many ways in which people can discern relations holding between widely scattered elements or objects in a scene. Some writings inspired by change blindness would seem to imply that human visual experience is akin to the experience of looking at the world through a long cardboard tube. However, people demonstrably have a capacity for visual perception of structure that does not fit with such a view (indeed, many of the tasks described earlier simply could not be performed to any reasonable degree by a creature limited in this way). Boolean map theory offers a description of the representational content of human visual experience that may conform more closely to the peculiar combination of strengths and weaknesses that characterizes human perceptual abilities. The theory entails massive parallelism and enormous flexibility but also some striking representational limitations.

Boolean Map Theory Offers a Potential Unification of Previous Disparate Theories

We have shown that Boolean map theory provides a unified framework in which to view target–target competition (Duncan, 1980a, 1980b), feature integration (Treisman & Gelade, 1980), object-based attention (Duncan, 1984), and change blindness (Pashler, 1988b; Rensink, 2000a, 2000b). What makes this set of phenomena an interesting group of ideas to try to unify is the fact that they are usually viewed in connection with opposing theoretical ideas. Above, we have discussed how an explicit distinction between selection and access explains the apparent contradiction between the parallel feature processing in feature-integration theory and the strong representational constraints that (we contend) govern conscious access to features. For another example, the difficulty of conjunction search is typically explained by assuming that features are not automatically bound together, whereas the same-object advantage is taken as evidence that features are automatically bound together into perceptual objects. Likewise, this puzzling inconsistency is due to the lack of an explicit distinction between selection and access. Attention is necessary for binding in the sense that feature values from different dimensions cannot be directly combined before they are exploited by the mechanism of selection. However, they can nonetheless be represented simultaneously in a consciously accessed Boolean map (see also Nieuwenstein, Chun, van der Lubbe, & Hooge, 2005, for an interesting report on how a selection–access distinction can shed light on attentional blink). Boolean map theory provides a unified account that reconciles these findings.

General Discussion

We turn now to some conjectures and discussions that, though not integral to Boolean map theory, illustrate the potential of the approach to raise new questions and shed new light on some old and new issues in vision and visual cognition.

Visual Working Memory

It has long been recognized that visual working memory is closely connected to visual perception (Kosslyn, 1980; Phillips, 1974). Strictly speaking, Boolean map theory does not necessarily predict anything about visual working memory per se, but it would seem natural if Boolean map theory shed light upon the limitations and character of visual working memory. It has long been noted that the term *memory* does not necessarily entail a mechanism whose sole—or even primary—function is to retain information per se (A. Allport, 1980). Any mechanism that can be used to represent information over time can be called memory, and one way to retain spatial information over a short period would be to maintain a Boolean map to select those locations.

The apparent linkage between visual spatial working memory and visual attention proposed by Awh and Jonides (2001) would seem very compatible to this conjecture. These investigators found enhanced processing of new visual signals arriving at locations being maintained in a spatial working memory task (Awh, Jonides, & Reuter-Lorenz, 1998). On the other hand, moving attention away from memorized locations impaired memory for those locations (Smyth, 1996; Smyth & Scholey, 1994). Furthermore, it has been reported that visual spatial memory loads significantly disrupt visual search (Han & Kim, 2004; Woodman & Luck, 2004). These observations all suggest that maintaining visual spatial memory is at least partly accomplished by directing visual attention in an appropriate fashion (Awh & Jonides, 2001). This finding is obviously congenial to Boolean map theory with its contention that the distribution of spatial attention is data for spatial analysis.

What about visual (as against spatial) working memory? Visual working memory as usually assessed must include not only spatial information but also featural information (Phillips, 1974). Are the contents of visual working memory limited to one Boolean map? One map would not be sufficient to maintain anything more than the spatial distribution of a single-feature value. Performance levels in visual working memory tasks often show that people can maintain somewhat, but not a great deal, more information than that (e.g., Stefurak & Boynton, 1986; Wheeler & Treisman, 2002), and thus, visual working memory cannot be limited to just one Boolean map. One possibility is that the content of a few Boolean maps might be retained.[16] This seems potentially consistent with the finding of Jiang, Olson, and Chun (2000), who showed that visual working memory is organized around spatial configuration and that this configuration is in some ways more primitive than feature information. For example, memory for a feature of one item is disrupted by a change in the location of other items, whereas a change in the features of other items has little effect on the memory of the location. Jiang and colleagues also speculated that items of the same color are probably represented together (see also Kanizsa, 1979). These observations would all be consistent with the idea that visual working memory is represented as a collection of several Boolean maps.

Object-Based Attention

Object-based attention has been an important topic in recent research on visual attention. One question that has been hotly debated is whether object-based attention is always mediated by spatial attention (as grouped array theory contends; see Kramer, Weber, & Watson, 1997) or not (as the spatially invariant account claims; see Vecera & Farah, 1994). We believe part of the ambiguity surrounding this question is due to the lack of an explicit distinction between selection and access. If the question is asked about access, Boolean map theory holds that all visual information (featural information, object identity, etc.) is obligatorily indexed by locations and that the person must also access these locations when the features are accessed. On this point, the current view is in agreement with grouped array theory. On the other hand, if the question is posed with respect to selection, a Boolean map can be created from various types of nonspatial cues, so, from this per-

---

[16] This statement is not in contradiction to the earlier statement that there is only one Boolean map at any instant. Here, we merely imply that visual working memory is organized in a format similar to that of the Boolean map. These "Boolean maps" do not provide conscious access.

spective, attention can be nonspatial. Plainly speaking, spatial locations are always part of what is selected, but they are not always part of the cue to select.

Cuing and Processing Optimization

The term *visual attention* is often used to encompass two potentially distinct concepts. One of these relates to selection of visual information for conscious access, which is the focus of Boolean map theory. The other concept is that one act of selection may influence subsequent selection (and processing) of other visual information in close spatial proximity (e.g., cuing improves performance, Posner, 1980, or more directly, perceiving a particular object improves the perception of subsequent objects in the same location, Kim & Cave, 1995)—something we term *processing optimization*.

Two examples may help clarify the difference between the Boolean map and processing optimization. First consider a cuing paradigm. An initial display is composed of one red dot and one green dot (both briefly presented), followed by a display showing two characters (one digit and one letter). Suppose the digit is usually in the location of red dot and the letter is usually in the location of green dot (with this reversed in a small percentage of trials, say, 10%). Assume further that the observer's task is to make some judgment about the digit (e.g., is it odd or even?). In this situation, it would be common to regard the performance advantage in digit responses when the digit follows the red dot (cued) rather than when it follows the green dot (miscued) as a measure of attention to the red dot. This would correspond well to the concept of processing optimization. The Boolean map, on the other hand, corresponds to the mere fact that the red dot is seen. Therefore, if such an experiment revealed no difference as a function of whether the digit was cued versus miscued, one might—following conventional practice—conclude that attention was not paid to the red dot. This would be correct only with respect to processing optimization. However, if the observer saw the red dot (e.g., if the observer could state where the red dot was), then, according to the present theory, a Boolean map must have been created to encompass this dot, and in that sense, attention must have been paid to it.

In the above example of cuing, processing optimization corresponds to the conventional concept of attention. On the other hand, in the attentional tracking paradigm, participants attempt to track a few balls from a group of other, usually physically identical balls. By the conventions of the attentional tracking literature, attention to the tracked balls is said to be reflected in an ability to discriminate those attended balls from the others. This corresponds to the Boolean map. Processing optimization, on the other hand, would have to be measured otherwise. Suppose 10 characters suddenly appear in the location of 10 balls (one in each): If the characters in the positions formerly occupied by balls that were being tracked are perceived better, this would reflect processing optimization.

Equating attention with the Boolean map (i.e., selection and access) seems to be in line with some—but by no means all—prior usage (e.g., Duncan, 1980a, 1980b, 1984; Intriligator & Cavanagh, 2001; Lu & Sperling, 1995; Pylyshyn & Storm, 1988; Treisman & Gelade, 1980; Ullman, 1984). On the other hand, a large literature on visual attention relating to the cuing paradigm (e.g., Downing, 1988; Eriksen & St. James, 1986; Gobell, Tseng, & Sperling, 2004; Posner, 1980; Posner, Snyder, & Davidson, 1980; Yeshurun & Carrasco, 1998) has focused on processing optimization.

Though these aspects of attention seem clearly separable, to our knowledge, their general distinction has not been made explicitly in the literature—potentially causing some confusion. For example, although it is obvious that observers can perceive the spatial relationship between two spatially separate objects without bringing in objects between them, it has been considered a nonobvious question to ask whether attention can be paid to two spatially separate objects (e.g., Awh & Pashler, 2000; Bichot, Cave, & Pashler, 1999; Hahn & Kramer, 1998; Kramer & Hahn, 1995; McMains & Somers, 2004). Clearly, selection (and access) is what is at issue in the first case, whereas processing optimization is the topic in the second case. Also, in cuing paradigms, the unattended target is often detected only slightly more slowly and/or less accurately than the attended target. However, in many discussions, attention has been claimed to be a prerequisite for access to any visual information whatsoever (e.g., Rensink, 2000a, 2000b). Again, it appears that attention can only mean processing optimization in the first case and selection (and access) in the second case.

The present article is devoted exclusively to developing an account of selection and access, and for present purposes, we merely note that there is an important conceptual distinction between selection–access and processing optimization. Uncovering the empirical differences between the two—and we suspect there may be several important differences relating to time course, spatial precision, and other attributes—is a topic for future discussion and investigation.

Eye Movements and Boolean Map Theory

Attention and eye movements have a close functional relationship (for reviews, see Hoffman, 1998; Rayner, 1998). For example, a shift of visual attention to the target of an upcoming saccade seems to be required before the saccade can commence (Deubel & Schneider, 1996; Hoffman & Subramaniam, 1995). Consequently, as Liversedge and Findlay (2000) pointed out, efforts to model visual search that disregard the pattern of eye movements occurring during these tasks are probably passing up a potentially useful source of empirical constraints (see also Findlay, 1997; Scialfa & Joffe, 1998; Williams & Reingold, 2001; Zelinsky, 1996.) For the same reason, exploring the pattern of eye movements occurring in the kinds of perception-of-structure tasks examined here might well shed light on the ideas proposed in the current article. One very basic question to be examined in that regard is whether eye fixations directly reflect the spatial distribution of the Boolean map hypothesized in this article. For example, when an observer attempts to select both objects in Figure 3, does the eye tend to fixate upon something like the centroid of the Boolean map (even when, as in the figure, this is not even part of the selected region)? If that or some other relatively direct relationship generally holds, then eye movements are likely to prove very useful in exploring and testing the current approach and some of the tasks examined here.

Perceptual Grouping

Boolean map theory may also pose a challenge to some widely assumed interpretations of classic grouping phenomena. In Figure

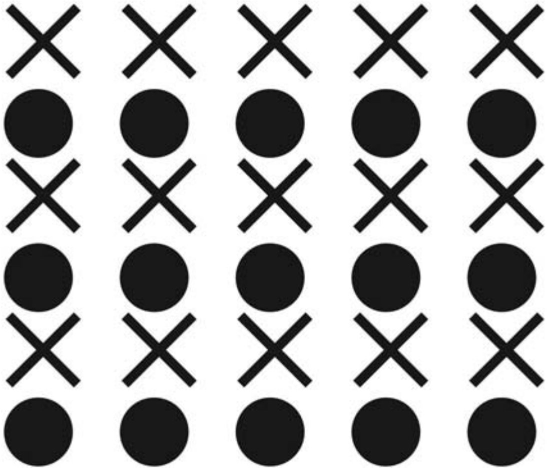


Figure 24. A probe of similarity grouping. Proximity grouping would favor a five-vertical-stripe organization, whereas similarity grouping—as traditionally understood—favors a six-horizontal-stripe organization. However, it seems that one can select either three horizontal stripes of balls or three horizontal stripes of crosses. Attempting to select both will obligatorily turn the organization into vertical stripes. We infer that similarity grouping operates by selecting one subset (e.g., all the crosses) to include in a Boolean map; proximity grouping is then applied to this map.

24, the display can be organized as columns according to Gestaltists' proximity grouping principle or as rows according to the similarity grouping principle. It is conventionally assumed that similarity grouping means this: Each item is linked together with any neighboring elements that are very similar to it (Kubovy, Holcombe, & Wagemans, 1998; Kubovy & Wagemans, 1995). However, is this really what happens? In the case of proximity grouping (vertical organization), all five columns do indeed seem to be simultaneously perceivable. However, in the case of similarity grouping, most people report that they can see the three rows of crosses or the three rows of balls—but not all six rows at the same time. If one tries to select all six, then the rows turn into columns. All of these observations can be made sense of in the following way. Suppose there is no general process of similarity grouping in the sense of various objects around a display becoming linked with other objects on the basis of various kinds of shared features. Instead, in line with Boolean map theory, suppose that similarity-based grouping is achievable only through a two-step process: (a) forming a Boolean map composed of one subset of stimuli of a certain type (e.g., all the crosses) and (b) applying proximity grouping to the elements that are represented in the map. Within the Boolean map of this display, items are closer to their horizontal neighbors than their vertical neighbors, and thus, one (but not both) of the two horizontal three-stripe arrangements should become perceptually available. We suspect that future research can probably uncover objective ways of testing the kind of hierarchical iterative grouping that we are hypothesizing.

Neural Underpinnings of the Boolean Map

In recent years, many researchers have been seeking to identify neural correlates or underpinnings of visual attention (e.g., Boynton, 2005; Bundesen, Habekost, & Kyllingsbaek, 2005; Gandhi, Heeger, & Boynton, 1999; Kastner, De Weerd, Desimone, & Ungerleider, 1998; O'Craven, Downing, & Kanwisher, 1999; Yantis et al., 2002). Boolean map theory is formulated at an algorithmic level, and thus, it does not directly imply any particular neural implementation. Nevertheless, some potential connections seem worth exploring.

The access constraints posited by Boolean map theory might fit quite well with a strikingly simple—one might even say crudely simple—interpretation of the idea of distinct "what?" and "where?" pathways (Mishkin, Ungerleider, & Macko, 1983), combined with a very simple idea about how binding across multiple specialized brain areas might be achieved. Imagine, for example, that there are topographic maps in the dorsal visual processing stream (chiefly in the parietal lobe) that represent the location of attended inputs, and suppose that a complex but stable pattern of activity there can potentially represent a complex configuration of multiple spatial locations. Coexisting with such a pattern of activity, suppose that in the ventral stream, there are numerous specific areas that represent different featural properties of the attended inputs (e.g., what color? what orientation? etc.) and that a stable pattern of activity in any one of these areas can represent one (but only one) specific value along that dimension. Given these premises, if one hypothesizes that a conscious visual percept consists of a set of stable states within both the dorsal and the ventral streams, the combination of the two would have precisely the representational capacity postulated for the Boolean map data structure—if one assumes that there are no mechanisms for binding beyond co-occurrence of these states. It seems at least conceivable, therefore, that the linkages between Boolean map theory and visual neurophysiology might turn out to be more straightforward than the linkages have been for other theories of visual attention. This point is offered merely as a conjecture; a detailed analysis of how Boolean map theory might be implemented in the brain is obviously a large task going far beyond the scope of the present article.

Concluding Comments

The present article has outlined a novel approach to visual attention and visual awareness. The basic two principles of Boolean map theory, stated in the theory section above, are not repeated here. In addition to offering a potential unification of some distinct (and, in a few cases, seemingly contradictory) aspects of visual attention, the present theory has several distinctive features, a number of which invite further investigation:

1. The theory proposes a representational format for visual conscious access (single-feature–multiple-locations data structure). Previous theorizing on limitations in access has shown that visual conscious access is capacity limited (Duncan, 1980a, 1980b), but (as far as we know) the present article offers the first effort to describe a testable set of representational constraints that might govern this access. The analysis seems strong and counterintuitive in its contention that only one feature per dimension can be accessed at one instant and that merely seeing a simple red–green object (see Figure 10a) requires sequential construction of two representations.

2. In this article, we have proposed a rather different conception of the principles of selection than has been proposed before, especially with regard to the exploitation of information from different dimensions and based on different feature values. Unlike the question of conscious access, the mechanisms of selection have been the subject of extensive theory, from feature-integration theory to the guided search model, and the topic has inspired a large body of research. The potential reinterpretation of these phenomena proposed here suggests many new lines of investigation.

3. We have advocated (and provided some examples of) the use of novel tasks designed to demand attention to structure. This may draw new empirical attention to a broad topic that has been neglected because of what, in our view, has been an overly narrow focus on visual search. As a concomitant benefit, such research may promote the development of richer connections between the study of visual attention and the rapidly developing field of information visualization (see Ware, 2000, for an excellent overview of this field, with a focus on connections to perceptual science).

References

Allport, A. (1980). Patterns and actions: Cognitive mechanisms are content-specific. In G. Claxton (Ed.), *Cognitive psychology: New directions* (pp. 26–64). London: Routledge & Kegan Paul.

Allport, D. A. (1971). Parallel encoding within and between elementary stimulus dimensions. *Perception & Psychophysics,* 10, 104–108.

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences, 5,* 119–126.

Awh, E., Jonides, J., & Reuter-Lorenz, P. A. (1998). Rehearsal in spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 780–790.

Awh, E., & Pashler, H. (2000). Evidence for split attentional foci. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 834–846.

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics, 55,* 485–496.

Barlow, H. B., & Reeves, B. C. (1979). Versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Research, 19,* 783–793.

Bauer, B., Jolicoeur, P., & Cowan, W. B. (1996). Visual search for colour targets that are or are not linearly separable from distractors. *Vision Research, 36,* 1439–1466.

Bertin, J. (1983). *Semiology of graphics.* Madison: University of Wisconsin Press.

Bichot, N. P., Cave, K. R., & Pashler, H. (1999). Visual selection mediated by location: Feature-based selection of noncontiguous locations. *Perception & Psychophysics, 61,* 403–423.

Boynton, G. (2005). Attention and visual perception. *Current Opinion in Neurobiology, 15,* 465–469.

Broadbent, D. A. (1958). *Perception and communication.* London: Pergamon Press.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97,* 523–547.

Bundesen, C., Habekost, T., & Kyllingsbaek, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review, 112,* 291–328.

Cavanagh, P., Labianca, A. T., & Thornton, I. M. (2001). Attention-based visual routines: Sprites. *Cognition, 80,* 47–60.

Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology, 22,* 225–271.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43,* 393–404.

Chong, S. C., & Treisman, A. (2005). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics, 67,* 1–13.

Davis, E. T., & Graham, N. (1981). Spatial frequency uncertainty effects in the detection of sinusoidal gratings. *Vision Research, 21,* 705–712.

Davis, E. T., Kramer, P., & Graham, N. (1983). Uncertainty about spatial frequency, spatial position, or contrast of visual patterns. *Perception & Psychophysics, 33,* 20–28.

Dennett, D. C. (1991). *Consciousness explained.* Boston: Little, Brown.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193–222.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research, 36,* 1827–1837.

Downing, C. J. (1988). Expectancy and visual spatial attention: Effects on perceptual quality. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 188–202.

Driver, J., & Baylis, G. C. (1989). Movement and visual attention: The spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 448–456.

Driver, J., & Vuilleumier, P. (2001). Perceptual awareness and its loss in unilateral neglect and extinction. *Cognition, 79,* 39–88.

Duncan, J. (1980a). Demonstration of capacity limitation. *Cognitive Psychology, 12,* 75–96.

Duncan, J. (1980b). The locus of interference in the perception of simultaneous stimuli. *Psychological Review, 87,* 272–300.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113,* 501–517.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433–458.

D'Zmura, M. (1991). Color in visual search. *Vision Research, 31,* 951–966.

Egeth, H. E. (1966). Parallel versus serial processes in multidimensional stimulus discrimination. *Perception & Psychophysics, 1,* 245–252.

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 32–39.

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics, 40,* 225–240.

Findlay, J. M. (1997). Saccade target selection during visual search. *Vision Research, 37,* 617–631.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1030–1044.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1993). Contingent attentional capture: A reply to Yantis (1993). *Journal of Experimental Psychology: Human Perception and Performance, 19,* 682–685.

Friedman-Hill, S., & Wolfe, J. M. (1995). Second-order parallel processing: Visual search for the odd item in a subset. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 531–551.

Gandhi, S. P., Heeger, D. J., & Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proceedings of the National Academy of Sciences, USA, 96,* 3314–3319.

Gobell, J. L., Tseng, C. H., & Sperling, G. (2004). The spatial distribution of visual attention. *Vision Research, 44,* 1273–1296.

Hahn, S., & Kramer, A. F. (1998). Further evidence for the division of attention among non-contiguous locations. *Visual Cognition, 5,* 217–256.

Han, S. H., & Kim, M. S. (2004). Visual search does not remain efficient when executive working memory is working. *Psychological Science, 15,* 623–628.

He, S., Cavanagh, P., & Intriligator, J. (1996, September 26). Attentional resolution and the locus of visual awareness. *Nature, 383,* 334–337.

Hoffman, J. E. (1979). A two-stage model of visual search. *Perception & Psychophysics, 25,* 319–327.

Hoffman, J. E. (1998). Visual attention and eye movements. In H. Pashler (Ed.), *Attention* (pp. 119–154). London: University College London Press.

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics, 57,* 787–795.

Howard, C., & Holcombe, A. O. (2007). *Progressively poorer perceptual precision and progressively greater perceptual lag: Tracking the changing features of one, two, and four objects.* Manuscript submitted for publication.

Huang, L., & Pashler, H. (2002). Symmetry detection and visual attention: A "binary-map" hypothesis. *Vision Research, 42,* 1421–1430.

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43,* 171–216.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20,* 1254–1259.

Jiang, Y. H., Chun, M. M., & Olson, I. R. (2004). Perceptual grouping in change detection. *Perception & Psychophysics, 66,* 446–453.

Jiang, Y. H., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 683–702.

Johnston, J. C., & Pashler, H. (1990). Close binding of identity and location in visual feature perception. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 843–856.

Julesz, B. (1981, March 12). Textons, the elements of texture perception, and their interactions. *Nature, 290,* 91–97.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. A. Davis (Eds.), *Varieties of attention* (pp. 29–62). New York: Academic Press.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24,* 175–219.

Kanizsa, G. (1979). *Organization in vision: Essays on gestalt perception.* New York: Praeger Publishers.

Kaptein, N. A., Theeuwes, J., & van der Heijden, A. H. C. (1995). Search for a conjunctively defined target can be selectively limited to a color-defined subset of elements. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1053–1069.

Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L. C. (1998, October 12). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science, 282,* 108–111.

Kawabata, N., Yamagami, K., & Noaki, M. (1978). Visual fixation points and depth perception. *Vision Research, 18,* 853–854.

Kim, M. S., & Cave, K. R. (1995). Spatial attention in visual search for features and feature conjunctions. *Psychological Science, 6,* 376–380.

Koenderink, J. J., & van Doorn, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America: Optics, Image Science, and Vision, 8*(A), 377–385.

Koshino, H., Warner, C. B., & Juola, J. F. (1992). Relative effectiveness of central, peripheral, and abrupt-onset cues in visual attention. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 45*(A), 609–631.

Kosslyn, S. M. (1980). *Image and mind.* Cambridge, MA: Harvard University Press.

Kramer, A. F., & Hahn, S. (1995). Splitting the beam: Distribution of attention over noncontiguous regions of the visual field. *Psychological Science, 6,* 381–386.

Kramer, A. F., Weber, T. A., & Watson, S. E. (1997). Object-based attentional selection—Grouped arrays or spatially invariant representations? Comment on Vecera and Farah (1994). *Journal of Experimental Psychology: General, 126,* 3–13.

Kubovy, M. (1981). *Psychology Survey,* Vol. 2—Connolly, K. *Contemporary Psychology, 26,* 471–472.

Kubovy, M. (1988). Should we resist the seductiveness of the space:time::vision:audition analogy? *Journal of Experimental Psychology: Human Perception and Performance, 14,* 318–320.

Kubovy, M., Holcombe, A. O., & Wagemans, J. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology, 35,* 71–98.

Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition, 80,* 97–126.

Kubovy, M., & Wagemans, J. (1995). Grouping by proximity and multistability in dot lattices: A quantitative gestalt theory. *Psychological Science, 6,* 225–234.

LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review, 96,* 101–124.

Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review, 4,* 501–506.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Science, 4,* 6–14.

Lu, Z. L., & Sperling, G. (1995). The functional architecture of human visual motion perception. *Vision Research, 35,* 2697–2722.

Magnussen, S., & Greenlee, M. W. (1997). Competition and sharing of processing resources in visual discrimination. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1603–1616.

Maruff, P., Danckert, J., Camplin, G., & Currie, J. (1999). Behavioral goals constrain the selection of visual information. *Psychological Science, 10,* 522–525.

McMains, S. A., & Somers, D. C. (2004). Multiple spotlights of attentional selection in human visual cortex. *Neuron, 42,* 677–686.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences, 6,* 414–417.

Moore, C. M., & Egeth, H. (1998). How does feature-based attention affect visual processing? *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1296–1310.

Morales, D., & Pashler, H. (1999, May 13). No role for colour in symmetry perception. *Nature, 399,* 115–116.

Navon, D. (1977). Forest before trees: Precedence of global features in visual perception. *Cognitive Psychology, 9,* 353–383.

Nieuwenstein, M. R., Chun, M. M., van der Lubbe, R. H. J., & Hooge, I. T. C. (2005). Delayed attentional engagement in the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1463–1475.

O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999, October 7). fMRI evidence for objects as the units of attentional selection. *Nature, 401,* 584–587.

O'Regan, J. K., & Noe, A. (2001). A sensorimotor account of vision and. visual consciousness. *Behavioral and Brain Sciences, 24,* 939–1031.

Palmer, S. E., & Hemenway, K. (1978). Orientation and symmetry: Effects of multiple, rotational, and near symmetries. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 691–702.

Pashler, H. (1988a). Cross-dimensional interaction and texture segregation. *Perception & Psychophysics, 43,* 307–318.

Pashler, H. (1988b). Familiarity and visual change detection. *Perception & Psychophysics, 44,* 369–378.

Pashler, H. (1998). *The psychology of attention.* Cambridge, MA: MIT Press.

Pashler, H., & Harris, C. R. (2001). Spontaneous allocation of visual attention: Dominant role of uniqueness. *Psychonomic Bulletin & Review, 8,* 747–752.

Peterson, M. A., de Gelder, B., Rapcsak, S. Z., Gerhardstein, P. C., & Bachoud-Levi, A. C. (2000). Object memory effects on figure assignment: Conscious object recognition is not necessary or sufficient. *Vision Research, 40,* 1549–1567.

Phillips, W. A. (1974). Distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16,* 283–290.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32,* 3–25.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109,* 160–174.

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition, 32,* 65–97.

Pylyshyn, Z. W. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual Cognition, 11,* 801–822.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3,* 1–19.

Quinlan, P. T. (2003). Visual feature integration theory: Past, present, and future. *Psychological Bulletin, 129,* 643–673.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422.

Remington, R. W., & Folk, C. L. (2001). A dissociation between attention and selection. *Psychological Science, 12,* 511–515.

Rensink, R. A. (2000a). The dynamic representation of scenes. *Visual Cognition, 7,* 17–42.

Rensink, R. A. (2000b). Seeing, sensing, and scrutinizing. *Vision Research, 40,* 1469–1487.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8,* 368–373.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (2000). On the failure to detect changes in scenes across brief interruptions. *Visual Cognition, 7,* 127–145.

Rock, I. (1983). *The logic of perception.* Cambridge, MA: MIT Press.

Rubin, E. (2001). Figure and ground. In S. Yantis (Ed.), *Visual perception: Essential readings* (pp. 225–229). Philadelphia: Psychology Press. (Original work published 1921)

Saiki, J. (2003). Feature binding in object-file representations of multiple moving items. *Journal of Vision, 3,* 6–21.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition, 80,* 1–46.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology, 38,* 259–290.

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition, 80,* 159–177.

Scholl, B. J., Pylyshyn, Z. W., & Franconeri, S. L. (2007). *The relationship between property-encoding and object-based attention: Evidence from multiple object tracking.* Manuscript submitted for publication.

Scialfa, C. T., & Joffe, K. M. (1998). Response times and eye movements in feature and conjunction search as a function of target eccentricity. *Perception & Psychophysics,* 1067–1082.

Shiffrin, R. M., & Gardner, G. T. (1972). Visual processing capacity and attentional control. *Journal of Experimental Psychology,* 93, 72–82.

Shih, S. I., & Sperling, G. (1996). Is there feature-based attentional selection in visual search? *Journal of Experimental Psychology: Human Perception and Performance, 22,* 758–779.

Slotnick, S. D., & Yantis, S. (2005). Common neural substrates for the control and effects of visual attention and perceptual bistability. *Cognitive Brain Research, 24,* 97–108.

Smyth, M. M. (1996). Interference with rehearsal in spatial working memory in the absence of eye movements. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49*(A), 940–949.

Smyth, M. M., & Scholey, K. A. (1994). Interference in immediate spatial memory. *Memory & Cognition, 22,* 1–13.

Stefurak, D. L., & Boynton, R. M. (1986). Independence of memory for categorically different colors and shapes. *Perception & Psychophysics, 39,* 164–174.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics, 51,* 599–606.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95,* 15–48.

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 459–478.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Tsal, Y., & Kolbet, L. (1985). Disambiguating ambiguous figures by selective attention. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 37*(A), 25–37.

Ullman, S. (1984). Visual routines. *Cognition, 18,* 97–159.

van der Helm, P. A., & Leeuwenberg, E. L. J. (1996). Goodness of visual regularities: A nontransformational approach. *Psychological Review, 103,* 429–456.

Vecera, S. P., & Farah, M. J. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology: General, 123,* 146–160.

Vecera, S. P., Flevaris, A. V., & Filapek, J. C. (2004). Exogenous spatial attention influences figure-ground assignment. *Psychological Science, 15,* 20–26.

Ware, C. (2000). *Information visualization: Perception for design.* San Francisco: Morgan Kaufmann.

Warner, C. B., Juola, J. F., & Koshino, H. (1990). Voluntary allocation versus automatic capture of visual attention. *Perception & Psychophysics, 48,* 243–251.

Watson, A. B., & Robson, J. G. (1981). Discrimination at threshold: Labeled detectors in human vision. *Vision Research, 21,* 1115–1122.

Watson, D. G., Maylor, E. A., & Bruce, L. A. M. (2005). The efficiency of feature-based subitization and counting. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1449–1462.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131,* 48–64.

Williams, D. E., & Reingold, E. M. (2001). Preattentive guidance of eye movements during triple conjunction search tasks: The effects of feature discriminability and saccadic amplitude. *Psychonomic Bulletin & Review, 8,* 476–488.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202–238.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 419–433.

Woodman, G. F., & Luck, S. J. (2004). Visual search is slowed when visuospatial working memory is occupied. *Psychonomic Bulletin & Review, 11,* 269–274.

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology, 24,* 295–340.

Yantis, S. (1993). Stimulus-driven attentional capture and attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 676–681.

Yantis, S., & Egeth, H. E. (1999). On the distinction between visual salience and stimulus-driven attentional capture. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 661–676.

Yantis, S., & Johnson, D. N. (1990). Mechanisms of attentional priority. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 812–825.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 601–621.

Yantis, S., Schwarzbach, J., Serences, J. T., Carlson, R. L., Steinmetz, M. A., Pekar, J. J., & Courtney, S. M. (2002). Transient neural activity in human parietal cortex during spatial attention shifts. *Nature Neuroscience, 5,* 995–1002.

Yeshurun, Y., & Carrasco, M. (1998, November 5). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature, 396,* 72–75.

Zelinsky, G. (1996). Using eye saccades to assess the selectivity of search movements. *Vision Research, 36,* 2177–2187.

(*Appendix follows*)

Appendix

Experiments

General Method

Participants

Undergraduate students from the University of California, San Diego, participated in this project. All participants had normal or corrected-to-normal vision. There were 9 participants in Experiment 1, 8 participants in Experiment 2, 18 participants in Experiment 3, 16 participants in Experiment 4, and 16 participants in Experiment 5.

Apparatus

Stimuli were presented on a 1,024 pixel × 768 pixel color monitor. Participants viewed the displays from a distance of about 60 cm and entered responses using the keyboard. The program was written in Microsoft Visual Basic 6.0 and was run in Microsoft Windows XP, using timing routines tested with the Blackbox Toolkit.

Procedure

Each trial began with a small white fixation cross presented for 400 ms in the center of the screen. After a short blank interval (400 ms), the stimuli were presented. Participants made the appropriate decision and responded by pressing one of two adjacent keys (*j* and *k*) with fingers of the right hand after all stimuli had been presented (in some experiments, there was more than one frame). In Experiments 1–2, the stimuli remained in the display until response, and participants were asked to respond as accurately and quickly as possible. In Experiments 3–5, the stimuli were masked (details below), and participants were asked to respond as accurately as possible (unspeeded response). A tone sounded to indicate whether the response was correct, and the next trial began 400 ms later.

Experiments 1–2: Repetition and Rotation of Color Patterns

Method

Sample displays in ABBA and ABCD conditions from Experiment 1 are shown above in Figure 7. In Experiment 1 (repetition experiment), two 4 × 4 color patterns were presented against a gray background. They were each 5.6 cm × 5.6 cm (each square 1.4 cm × 1.4 cm). The distance between their centers was 8.4 cm (gap = 2.8 cm). We used the four colors red, green, blue, and yellow. The two patterns might be identical or different, each with a probability of 50%. Participants responded to this by pressing two keys (*j* for different, *k* for identical). When the patterns were identical, the four colors were each used four times. When they were different, two positions with different colors were randomly chosen from the display. There were two types of display. In the ABBA condition, the two changed squares switched color with each other. In the ABCD condition, the two changed squares were each changed to a new color. For example (see Figure 7), suppose that two positions were chosen and that they were red and green. In the ABBA condition, the red square was changed to green, and the green square was changed to red. In the ABCD condition, the red square was changed to yellow, and the green square was changed to blue.

In Experiment 2 (rotation experiment), everything else was the same as Experiment 1, except that the right pattern was rotated 90° clockwise or anticlockwise and the participants judged if the patterns were identical before the rotation. The direction of rotation was constant across a whole block and alternated from block to block. The direction of rotation was explicitly given before each block.

In both the repetition experiment and the rotation experiment, participants finished five blocks (one block had 100 trials in the repetition experiment and 50 trials in the rotation experiment.) The first block was excluded from data analysis as practice.

Results

The logic of this method is as follows: In the ABCD condition, the Boolean map of any color will suffice to reveal that the two patterns are different. In the ABBA condition, only the Boolean maps of two colors will reveal that the two patterns are different. Therefore, if participants are serially scanning from one color to another color, then they always have to check only one color to detect a difference in the ABCD condition, whereas, in the ABBA condition, they sometimes have to check two or even three colors to detect a difference. Thus, response times should be longer in the ABBA condition. Otherwise, if participants are doing parallel comparison of the entire pattern, the two conditions should contain equivalent amounts of difference for the ABBA and ABCD conditions.

The response time of the ABBA condition was longer than that of the ABCD condition in both the repetition experiment (2,100 ms vs. 1,905 ms, difference = 195 ms), $F(1, 8) = 11.96$, $p < .01$, and the rotation experiment (3,463 ms vs. 2,659 ms, difference = 804 ms), $F(1, 7) = 16.76$, $p < .01$. Consistently, the error rates of the ABBA condition were also higher in both the repetition experiment (7.7% vs. 5.3%), $F(1, 8) = 4.44$, $p < .025$, and the rotation experiment (14.4% vs. 5.5%), $F(1, 7) = 11.74$, $p < .02$. Therefore, the results of Experiments 1–2 fit the claim of Boolean mapping theory that the comparison is being done by a serial scanning strategy from color to color rather than by a parallel comparison of the entire patterns.

Experiments 3–4: Simultaneous Access to Two Color Values or Two Location Values

Method

Sample displays and procedures are illustrated above in Figure 10. In Experiment 3 (see Figure 10a), a wheel (diameter = 4.2 cm) was divided into four quadrants. The top left and bottom right quadrants shared one color, and the bottom left and top right quadrants shared one color. Four colors (red, green, blue, yellow) were used in Experiment 3, and in each trial, two of the four were randomly picked to constitute the wheel, and one of the four was randomly chosen to be the probe color, which was presented as a square (0.78 cm × 0.78 cm) in the center of the screen. The two colors of the wheel were presented either successively (interframe interval = 700 ms) or simultaneously. In both cases, the displays were rapidly masked (duration of mask = 200 ms) after a very brief exposure. The duration of stimuli exposure was adjusted for each participant to achieve a moderate performance (ranging from 23 ms to 61 ms, with an average of 42 ms). The probe color was presented 700 ms after the offset of the last frame (i.e., the only frame in the simultaneous condition and the second frame in the successive condition) and remained present until response. In Experiment 3, participants were required to perceive the two colors in the wheel and later to determine if the probe color was one of them.

In Experiment 4 (see Figure 10b), yellow squares could be presented in four locations (four quadrants) and were divided into two pairs (one pair was top left and bottom right; the other pair was bottom left and top right). In each trial, one square was randomly picked from each pair to constitute the stimulus, and one of the four was randomly chosen to be the probe location. Squares measured 0.78 cm × 0.78 cm each and were 0.91 cm off the center of the display both vertically and horizontally. The two squares were presented either successively (interframe interval = 700 ms) or simultaneously. In both cases, the displays were rapidly masked (duration of mask = 200 ms) after a very brief exposure. The duration of stimuli exposure was adjusted for each participant to achieve a moderate performance (ranging from 12 ms to 50 ms, with an average of 30 ms). The probe location was presented 700 ms after the offset of the last frame (i.e., the only frame in the simultaneous condition and the second frame in the successive condition) and remained present until response. In Experiment 4, participants were required to perceive the locations of the two squares and later to determine if the probe location was one of them.

In both Experiments 3 and 4, participants performed 10 blocks of 50 trials. In each block, the trials switched between simultaneous and successive presentations. The first block was excluded from data analysis as practice.

Results

In Experiment 3, the average accuracy in the successive condition was significantly better than that in the simultaneous condition (successive condition 0.734 vs. simultaneous condition 0.665), $F(1, 17) = 41.23$, $p < .0001$. In Experiment 4, the average accuracy in the successive condition was slightly worse than in the simultaneous condition (successive condition 0.743 vs. simultaneous condition 0.770), $F(1, 15) = 3.93$, $p < .1$. The interaction was significant, $F(1, 32) = 31.17$, $p < .0001$. These results suggest that there is a severe difficulty in accessing two colors simultaneously, whereas there is no such difficulty in accessing two locations simultaneously.

A model was constructed to ask how sequential the color access really was in Experiment 3. Assume that the performance is limited by the strength of the early sensory signal and also the restriction of visual access. To simplify the question, we assumed a dichotomy whereby, in a certain percentage of trials (p), the signal is strong enough to allow conscious access, and in the rest ($1 - p$), it is not. We also assumed that the strength of early sensory signal is independent for each color. In the successive condition, each color is perceived in p trials; therefore, the accuracy should be $0.5 + p/2$. In the simultaneous condition, the early sensory signal allows zero colors to be perceived in $(1 - p)(1 - p)$ trials, one color to be perceived in $2p(1 - p)$ trials, and two colors to be perceived in pp trials. The restriction of conscious access applies to the latest condition (two colors), so only one color can be perceived in those trials. Taken together, in the simultaneous condition, one color can be perceived in $2p - pp$ trials; therefore, each of the two colors can be perceived in $p - pp/2$ trials. That corresponds to an accuracy of $0.5 + (p/2) - (pp/4)$. The best fitting value of p is 0.456. Given that, the predicted accuracy for simultaneous and successive conditions is 0.728 and 0.676, respectively, and very closely resembles the actual data of 0.734 and 0.665. In short, the significant but not enormous difference between performance in the simultaneous and successive conditions is consistent with strictly sequential access to colors.

Experiment 5: Time Course of Feature Search and Conjunction Search

Method

In Experiment 5, participants searched for a red vertical target either from a uniform array of yellowish red vertical bars (feature search task) or an array of red horizontal bars and green vertical bars (conjunction search task). In each display, 32 items (including a target) were placed in arrays of nine columns and four rows, with no item placed in the center column (i.e., two 4 × 4 arrays with a gap in the center). The distances between centers of items were 1.30 cm both vertically and horizontally. Each bar was 0.78 cm long and 0.16 cm wide. In each trial, there was always one target. Participants decided whether it was on the left side (left four columns) or right side (right four columns) of the display and responded. The display was presented for 100 ms, 200 ms, or 400 ms before being masked. In Experiment 5, participants performed 12 blocks of 80 trials. The first two blocks were excluded from data analysis as practice. The blocks alternated between feature search blocks and conjunction search blocks, and the starting block was balanced across participants.

Results

In Experiment 5 (see Figure 22, above), the performance of conjunction search was inferior to feature search at short exposure but superior at long exposure; the interaction was significant, $F(1, 15) = 16.39$, $p < .002$. This distinctive pattern suggests that conjunction search relies upon a dimension-to-dimension subset search strategy, not a top-down salience map (see text for details).

Received November 3, 2005
Revision received December 28, 2006
Accepted January 2, 2007 ■