measuring the limits of your visual system
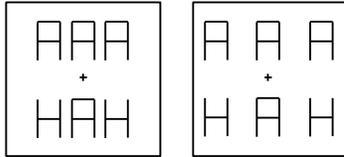
resource-consuming tasks from visual cognition studies

Cueing Benefits: A location is cued with a flash or arrow, responses are faster or more accurate to probe at that location.
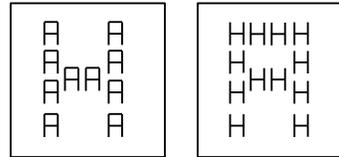
Pop-out vs. Slow Visual Search: Finding the bright object at left is immediate, finding the backward C at right takes time.
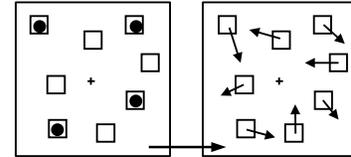
Selecting Multiple Locations: Finding the grey vertical at right takes time, but we can select a limited number of hint locations
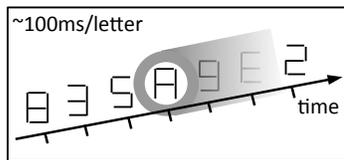
Flanker Interference: Deciding if center letter is A or H is slower when flankers are incompatible (bottom). Greater spacing helps.
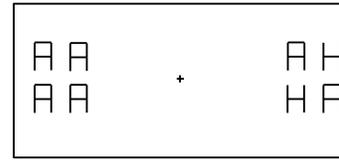
Global Interference: Deciding if smaller letter is A or H is slower when large letter is incompatible. Smaller effect in opposite direction.
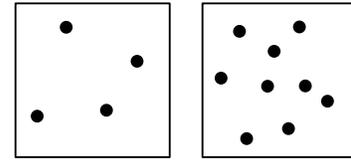
Tracking Multiple Objects: We can track a few cued moving objects at fast speeds, and up to 8 at slower speeds.

~100ms/letter

time

Attentional Blink: While reporting letters in a rapidly changing number stream, processing 'blinks' after a target: 'A' is identified, 'E' is missed.

Crowding: Fixate. Letters at left are clear. Letters at right seem to have floating or interchanging identities.
[adapted from Pelli, Palomares, & Majaj, 2004]

Fast vs. Slow Enumeration: We can instantly count up to 4 objects. For more objects, counting is very slow (though we can quickly estimate).

Averaging: We can quickly judge average size in right display. Remaining fixated on right display, note average orientation on left.

Illusory Conjunctions: Forcing attention to spread widely with number judgment, percept of shape features can miscombine.

Slow Tracing: Keeping your eyes fixed, is the dot at left inside or outside the shape? At right, are the two dots on the same line?

Inattentional Blindness: Count wall bounces of white objects. People miss appearance of new black objects, red objects, gorillas.

Object Substitution: An array of shapes is briefly flashed, we can't report a shape masked by dots, unless its position is pre-cued.

Spatial Relationship Perception: Is surprisingly slow. Find the two pairs in each display with unique relationships.

Ventral Competition For Identification

Dorsal Competition For Selection

But what is the resource?

CHAPTER

10

The Nature and Status of Visual Resources

Steven L. Franconeri

Abstract

Across many types of tasks, our ability to process multiple objects or locations at once is limited by a finite processing resource. This chapter describes 15 classic examples of such resources limits. The chapter then reviews evidence suggesting that this resource primarily reflects competition for representation, across two types of representation. First, limits on the identification of objects may reflect competition within networks that represent object identity (the ventral visual stream). Second, limits on the selection of multiple locations may reflect competition between selected locations within representations of visual space (the dorsal visual stream). This definition of visual resources provides a parsimonious explanation for many effects in the visual cognition literature, and it makes concrete predictions about manipulations that should affect performance across a wide variety of visual tasks.

Key Words: visual attention, selection, resources, competition, crowding, surround inhibition, object tracking, visual search, subitizing

The goal of research in visual cognition is to explain how our visual system takes a stream of information from tens of millions of constantly changing visual "pixels" and produces a coherent visual experience that can lead to useful action in the world. This is a daunting problem, because it is difficult to comprehend the huge space of possible ways that the visual system might achieve this result. One way to reduce the scope of any problem is to break it into smaller pieces, in this case by reducing the visual system to more constrained systems. These subsystems might process certain kinds of stimuli, such as color, texture, or motion. Or they might perform a certain type of processing, such as edge detection, which is any process that seeks contrast among different colors, textures, or motion directions.

Among the constrained problems studied by vision researchers, *visual attention* is an odd class because it might be defined as "processing that fails when you give it too much to do at once." These failures, illustrated in Figure 10.1, seem to show that as people are asked to deal with a larger number of objects, there is a limited pool of visual "resources" that becomes depleted, leading to slower response time or lower accuracy. The chapter argues for a common root of these limited resources: competition for representation, either within a representational space of object identity (the ventral visual system), or a representational space that stores currently selected locations (a dorsal map of visual space).

The second section reviews the many examples of processing limitations depicted in Figure 10.1. The third section isolates the examples that reflect visual resource limitations (as opposed to, e.g., limits on higher level decision processes). The fourth section reviews evidence that visual processing can be limited by competition for representation, and the fifth and sixth sections argue that this resource can
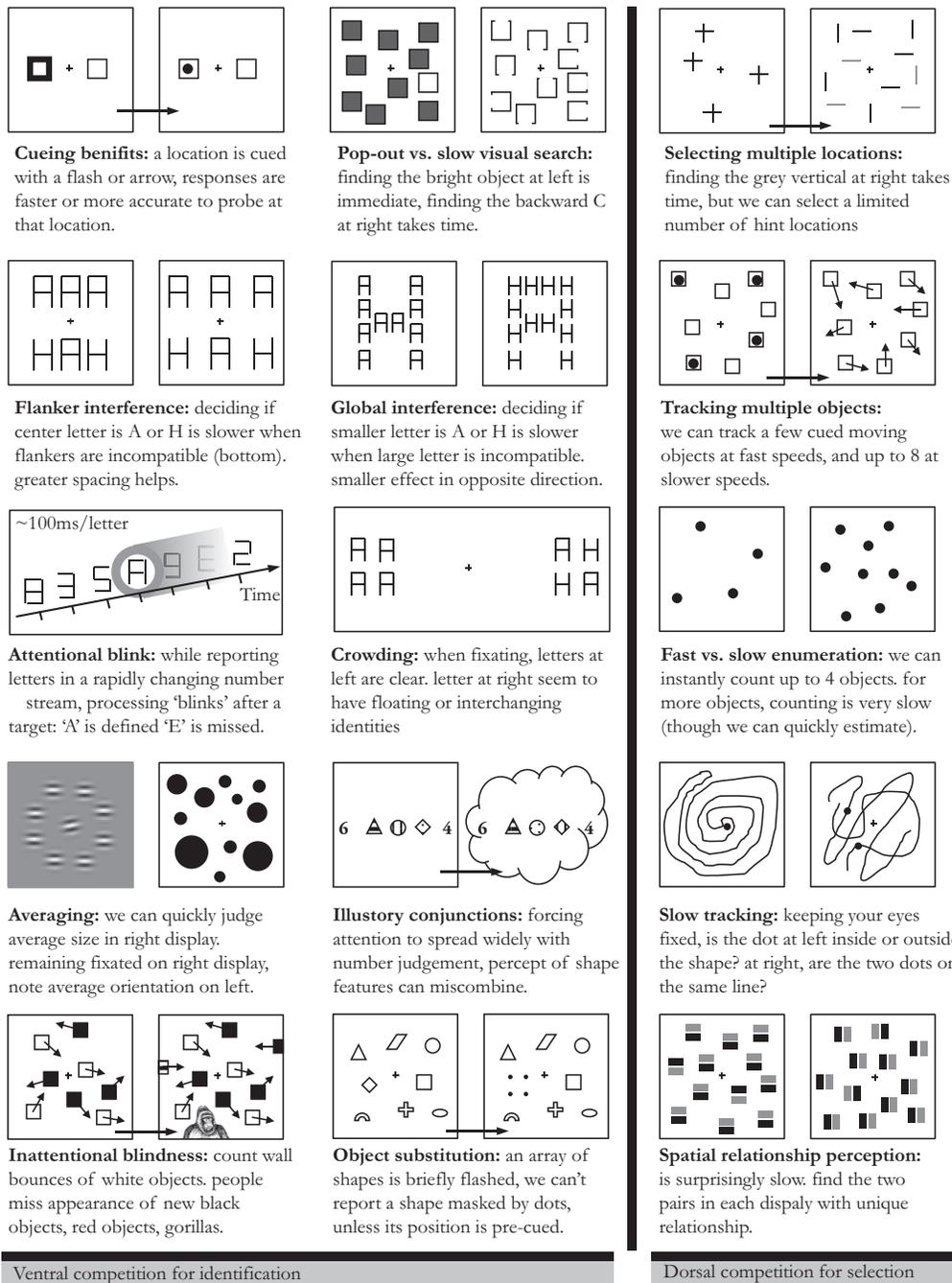
Cueing benifits: a location is cued with a flash or arrow, responses are faster or more accurate to probe at that location.

Pop-out vs. slow visual search: finding the bright object at left is immediate, finding the backward C at right takes time.

Selecting multiple locations: finding the grey vertical at right takes time, but we can select a limited number of hint locations

Flanker interference: deciding if center letter is A or H is slower when flankers are incompatible (bottom). greater spacing helps.

Global interference: deciding if smaller letter is A or H is slower when large letter is incompatible. smaller effect in opposite direction.

Tracking multiple objects: we can track a few cued moving objects at fast speeds, and up to 8 at slower speeds.

Attentional blink: while reporting letters in a rapidly changing number stream, processing 'blinks' after a target: 'A' is defined 'E' is missed.

Crowding: when fixating, letters at left are clear. letter at right seem to have floating or interchanging identities

Fast vs. slow enumeration: we can instantly count up to 4 objects. for more objects, counting is very slow (though we can quickly estimate).

Averaging: we can quickly judge average size in right display. remaining fixated on right display, note average orientation on left.

Illustory conjunctions: forcing attention to spread widely with number judgement, percept of shape features can miscombine.

Slow tracking: keeping your eyes fixed, is the dot at left inside or outside the shape? at right, are the two dots on the same line?

Inattentional blindness: count wall bounces of white objects. people miss appearance of new black objects, red objects, gorillas.

Object substitution: an array of shapes is briefly flashed, we can't report a shape masked by dots, unless its position is pre-cued.

Spatial relationship perception: is surprisingly slow. find the two pairs in each dispaly with unique relationship.

Ventral competition for identification

Dorsal competition for selection

Figure 10.1 Fifteen examples of visual tasks that have been taken to reflect limits on visual resources. Examples on the left are argued to reflect competition for representation in the ventral visual stream and, on the right, the dorsal visual stream. (The "crowding" example is adapted from Pelli, Palomares, & Majaj, 2004.)

explain performance limits in ventral object identification (left side of Fig. 10.1) and dorsal object selection (right side of Fig. 10.1). The final section offers a broader picture of how these limits might affect everyday perception.

Fifteen Examples of Visual Resource Limits

Figure 10.1 lists examples of tasks that reveal limitations in our ability to process visual information. All of them show that when we ask too much of our visual system, response times lengthen or accuracy

falls. Some of these limitations may be illusory, but all have historically served as examples.

Limits on Identification of Objects

Cueing effects can occur when observers are asked to make a fast response when a probe dot appears or to identify a letter that appears. There are two (or more) potential locations for this probe, and performance is faster or more accurate if the target's location is cued beforehand, relative to when another location is cued (Posner, 1980). The cue seems to allow faster processing when an observer focuses on a single location.

Limits on *visual search* can occur when observers are asked to find a target object in a field of distractor objects (Wolfe, 2007). For some versions of the task, response times are fast no matter how many distractor objects are added to the display. This might include searches for a bright object among dark objects, searches for a red object among green objects, or searches for a moving object among static objects. But for other versions of the task, response times increase as more distractors are added. These include searches for a target letter among other letters, or a search for a conjunction of two features (e.g., a gray vertical bar) among distractors that each carry one of the target's features (e.g., gray horizontal bars and black vertical bars, see the "Multiple location selection" example lower in the figure). For these searches, we appear to be limited in the number of objects that we can process at once.

Flanker interference effects occur when known target objects are flanked by nearby objects that carry an incompatible response. If an observer's task is to determine whether a center letter is an A or an H, and the center letter is an A, responses are slower when the flankers are H's than when the flankers are also A's. Moving the letters farther away, or making them less featurally similar (e.g., different colors), reduces this interference effect. Something about the addition of these incompatible letters limits the observer's ability to respond quickly (see Chajut, Schupak, & Algom, 2009; Mordkoff, 1996). In another type of interference effect labeled *global interference*, the identity of a larger object can interfere with recognition of smaller objects that comprise it. In the figure example, observers are asked to detect whether the small letters are A's or H's. Similar to the flanker effect, incompatible global letters slow processing of local letters (Navon, 1977).

Attentional blink effects seem like an effort by the visual system to prevent flanker interference effects across time instead of space. When asked to identify two letters in a rapidly changing stream of digits (about 100ms/character), we can identify the first letter, but we are reduced in our ability to identify the second one. The typical conclusion is that in order to isolate the first digit, the object identification system must briefly "blink" in order to shut out interfering input from the subsequent "flanking" characters (Chun & Potter, 1995; Raymond, Shapiro, & Arnell, 1992).

The presence of too many object identities can even limit performance when observers are given ample time to inspect a display. In *crowding* effects, objects placed too close together in the visual periphery become difficult or impossible to identify. In Figure 10.1, fixate at the center cross and notice that the objects at the right have strangely jumbled identities. You can tell that the objects are roughly A's and H's, but the identities seem to float among the four letters (this example was adapted from Pelli, Palomares, & Majaj, 2004). This limit is not due to poor visual acuity in the periphery. When the object identities are homogeneous as on the left side, the letters are easy to distinguish. Instead, the important limiting factor is the number of distinct letter types that must be identified.

Interestingly, in the crowding example information about the letters is not lost, but instead the identities seem to become mixed together. Findings from tasks that examine visual *averaging* effects show that even if there is a limit to how well the information can be organized, the information can still be used to produce a summary of the otherwise jumbled features. In the averaging example, fixate your eyes on the cross in the right box and try to determine the orientation of the center patch in the left box. Even if you cannot tell the orientation of an individual patch, you may be able to judge the average orientation of the group as a whole (a slight counterclockwise tilt). Results from a similar task suggest that this average orientation is available from a crowded group, even if the identity of the center patch alone cannot be recovered independently (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). Other results suggest that a similar average representation is available even when objects are not crowded. If observers broadly select the group of circles at right, they are able to determine the average size of a set of circles (Ariely, 2001; Chong & Treisman, 2005; but see Myczek & Simons 2009 for evidence that this ability might instead be due to sampling of one or more circles). Later in the chapter this ability to extract average representations from crowded

objects will prove to be an interesting qualification to the limits implied by crowding.

The jumble of features created by crowding may have the same roots as a phenomenon called the *illusory conjunction* effect. Even when a display is not crowded, if observers are convinced to broadly select multiple objects (in the case in the figure, by requiring a response related to the numbers at each edge of the display), and the display is briefly flashed, the colors of the center shapes can become jumbled. A display containing a red triangle, green circle, and blue diamond might be misreported as containing a red triangle, blue circle, and green diamond (Treisman & Schmidt, 1982).

In the examples listed so far, limitations on visual processing appear to slow down responses, impair accuracy, or jumble the features of objects. But these limits can cause even more striking effects, where objects are never seen at all. In the phenomenon of *inattentional blindness*, engaging an observer in a demanding task, such as identifying line lengths in a quickly flashed display (Mack & Rock, 1998), counting the bounces of a set of moving shapes (Most, Scholl, Clifford, & Simons, 2005), or counting the bounces of a ball by a team of basketball players (Neisser & Becklen, 1975; Simons & Chabris, 1999), can cause observers to miss salient events right in front of their eyes. These events include novel shapes or colors (Mack & Rock, 1998; Most et al., 2005), a change in an ongoing action (Neisser & Becklen, 1975), or even a gorilla walking across the scene (Simons & Chabris, 1999). But if an observer is allowed to watch displays without focusing resources on the secondary task, they easily see the salient events (see Most et al., 2005, for review).

Object substitution masking effects rely on an even more direct manipulation that prevents an object from reaching awareness. If a set of objects are briefly displayed but then one is quickly replaced by a set of dots, the replaced object usually cannot be identified, even though the dots do not directly overlap the object's location (Enns & Di Lollo, 1997). However, if the observer is told where the object will appear before the trial so that he or she can focus on only one location, the observer easily sees the object.

Limits on Object Selection

Another set of tasks, shown on the right side of Figure 10.1, reveal limits on how we select objects and their locations over time, even when object identification is no longer necessary. Some reveal a limit on *multiple location selection*. In a typical task, observers complete a difficult visual search but are first precued to a set of potential target locations. The target will always appear at one of these locations, so it is beneficial to select them. Performance suggests that at least five (Burkell & Pylyshyn, 1997) and up to eight (Franconeri, Alvarez, & Enns, 2007) locations can be selected simultaneously, but no more. *Multiple object tracking* adds a second requirement. A set of objects are precued, and observers are asked to select them, but now the objects also move (Pylyshyn & Storm, 1988). This task is similar to the street magician's game of placing a valuable object under one of several quickly moving cups. With carefully constructed displays, observers can simultaneously track up to eight moving objects in this task (Alvarez & Franconeri, 2007).

In another limitation, there are strong effects of the number of objects when we are required to *enumerate* objects. There is little cost in response time when producing fast counts of sets of four or fewer objects, but when more than four objects are present, the number is not immediately available. Instead, response times suggest that for larger collections we need to count each object one at a time or settle for a rough estimate of the number (for review, see Dehaene & Changeux, 1993; Gallistel & Gelman, 1992). This limit of about four has been linked to the number of objects that we can simultaneously attend at once (see Trick & Pylyshyn, 1994, for review).

Another example reveals limits on selecting the location of an entire complex shape at once, which instead seems to require *tracing* an imaginary movement through the shape over time. Keeping your eyes fixed, determine whether the center dot is located inside or outside of the shape (Jolicoeur, Ullman, & Mackay, 1986; McCormick & Jolicoeur, 1991, 1994; Ullman, 1984; for similar examples). You may feel as though you trace a position through the maze (or "spread attention" through the maze; Houtkamp, Sprekreijse, & Roelfsma, 2003) and then see whether the dot can exit the shape. The feeling is similar for the second task, where you must decide whether the two dots are on the same or different lines. Making these decisions takes more time for longer or more complex lines.

Determining the relative *spatial relationships* among even objects is a highly resource-demanding process. When asked to find a pair of objects in a given configuration in a visual search display, adding additional pairs to the display slows response times significantly (Logan, 1994, 1995). A review

of visual search results shows spatial relationship searches to be among the more robustly difficult (Wolfe, 1998). Spatial relationship judgments are even time consuming with very small numbers of objects in a display—when deciding the relationship between two letters, adding just one additional letter can slow performance (Carlson & Logan, 2001).

Why Is Processing Limited?

All of the tasks mentioned so far reveal that visual processing becomes slower or less accurate when observers are asked to process too much information or too many objects. Why? A typical answer is that the visual system has only a finite *capacity* for processing, and as limited processing *resources* are spread more thinly with increased task *load*, speed and accuracy must be sacrificed (e.g., Broadbent, 1958; Kahneman, 1973; Norman & Bobrow, 1975). Words like capacity, resources, and load identify phenomena demand explanation (but see Navon, 1984, for a critique of the concept of "processing resources").

The challenge is to isolate a more concrete and useful picture of the limiting factor. Here is a good test of any possible definition: Given an omnipotent set of tools to change the mind and brain, could you predict what you would alter within the visual system to increase the capacity or resource, or reduce the limits or load? When speaking only of processing "resources," it's not clear what the answer should be. The next section discusses a more concrete proposal that can provide an answer to this question.

Before attempting to identify these roots in the examples mentioned earlier, it is important to acknowledge that many of the examples in Figure 10.1 may not reflect resource limitations after all (see Luck & Vecera, 2002, for additional discussion). First, just because a task feels difficult, takes time to complete, or is not performed with high accuracy, does not mean that more "resources" would necessarily help. Indeed, resource-limited processing can be contrasted with *data-limited* processing, where performance is capped by properties like the quality of incoming information (Norman & Bobrow, 1975). When processing is data limited, lowering the "task load," for example, asking observers to process fewer objects at once, does not help. As an example, imagine giving an observer a fast sequence of five letters that have been degraded with visual noise, and asking them to repeat the letters verbally. If recall averaged four letters (80% accuracy), it would be tempting to conclude that performance

reflected limited processing resources. But it is also likely that performance was data limited, by a letter that happened to be degraded in such a way that made it impossible to discern (say, a barely visible bottom section of a B made it confusable with a P). The critical test would be to manipulate the number of letters processed at once. If performance is data limited, then the overall score should not depend on the number of letters, but instead on the average performance combined across performance for each letter by itself. That is, it should depend solely on how many of those tricky B's are in the display. In contrast, resource-limited processing predicts that processing more letters at once, per se, should lower performance.

One important data limitation is the position of the eyes. Because visual resolution is best at the fovea, dropping off rapidly in the periphery, the position of the fovea represents a critical potential data limit. For example, in a visual search task, adding more objects slows responses. But adding more objects also increases the amount of space needed for objects in a display, usually pushing the objects farther from fixation. Thus, less accurate or slower processing of larger sets of objects might be due to lower quality of information for those objects, and performance for even one solitary object might also be lower at these more distant locations. The need to move the eyes in a task can even create set size effects that mimic a resource limit (Maioli, Beaglio, Siri, Sosta, & Cappa, 2001).

Another factor that can mimic a resource limit, but could be considered a data limit, is called decision noise. Consider the increase of response times with more distractors in visual search tasks, which is typically taken to reflect processing limitations. But these costs might occur even if every object in the display were identified simultaneously and independently. Because there is always noise in the visual system, there is always a chance that any one object will be recognized incorrectly (e.g., a distractor being mistaken as a target). Adding more objects to a display would therefore multiply the chances of at least one distractor being mistaken for the target, because there is a greater chance of at least one distractor having a sufficiently noisy representation (Duncan, 1980; Palmer, 1994).

One way to distinguish these decision limits from other resource limits is by presenting search objects independently over time instead of simultaneously. Decision limitations predict the same increasing difficulty with more distractors, because the same maximally noisy distractor representation

should be present within N distractors presented either independently or simultaneously (Duncan, 1980; see Pashler, 1998 for discussion). Resource limitations predict that independent presentation will eliminate the costs from additional objects, resulting in ceiling performance. This manipulation has shown that in many visual search tasks the additional distractor costs claimed to reflect resource limits could actually be better explained by decision noise (Huang & Pashler, 2005; Palmer, 1994, 1995). Note that the mechanisms for reducing the effects of decision noise, reducing the number of total objects inspected, or allowing more time to accumulate more information about each object could be considered resource limited. But it seems more appropriate to think of the visual system making mistakes on occasion, and when asking for a greater number of objects in a shorter amount of time, the odds of at least one troublesome mistake grow, even though processing of any individual object is no less thorough.

There are also strategic factors that can create illusory performance limitations. For example, in cueing tasks, detection or discrimination performance is better for cued locations, suggesting that processing was faster or more accurate when a limited resource could be selectively applied to the cued location. But cueing effects have also been explained by a set of strategy changes, such as altering the criterion for responding to information at cued locations, which are usually more likely to contain the target, or more heavily weighting information from cued locations. Some have even concluded that no cueing data exist that demonstrate resource limits outside of these strategic biases (Shiu & Pashler, 1994; Pashler, 1998; but see Luck & Thomas, 1999).

Finally, it should be noted that the limits described here are not the only limiting factors within visual cognition. First, there are limits not just in online perception but in short-term memory for object identities and locations. Storing visual information is limited by memory storage capacity, which appears to be limited to about four objects in many cases (Luck & Vogel, 1997; Zhang & Luck, 2008; but see Hollingworth, 2006). This might also constitute competition for representation, though in a different medium (e.g., feature space) than the examples listed in Figure 10.1. There may also be a limit to the process that compares information in memory to the information in the present view (Mitroff, Simons, & Levin, 2004). Even when memory limits are removed from change detection

tasks by presenting both displays simultaneously, finding changes is slow and difficult, suggesting that the comparison process is highly limited (Scott-Brown, Baker, & Orbach, 2001). There are limits at processing levels beyond vision, such as response selection (for review, see Pashler, 1998) that may limit abilities to complete tasks involving multiple objects, even when the actual processing of those objects does not require visual "resources."

What does this list of potentially illusory resource limitations mean for the examples in Figure 10.1? It is controversial whether cueing reflects resource limitations at all. Many search tasks also seem suspect, though there is general agreement that many types of visual search are resource limited (Huang & Pashler, 2005; Palmer, 1994; Wolfe et al., 2008). But all of the other limits found using the tasks listed in Figure 10.1 seem fairly immune from these critiques. So for the remaining examples, when observers are asked to process too much information at once, what is the resource that causes responses to slow, accuracy to drop, or awareness to fail? A general principle that seems to fit many examples is competition for representation within the visual system.

Competition for Representation

The first case of competition is for object recognition. Object recognition is handled primarily by the ventral stream of the visual system. This stream starts in primary visual cortex (V1), where networks filter incoming information by relatively simple properties like orientation or contrast polarity. Processing moves to subsequent areas that process more complex features (e.g., V4). The stream ends in inferior temporal (IT) cortex, where more sophisticated features such as shape are processed (see Grill-Spector & Malach, 2004 for review). In addition to increasing complexity of processing, moving farther along the ventral stream also greatly increases the area of the visual world that any cell responds to. In V1 this "receptive field" is less than 1 degree wide, but it grows to occupy large portions of the visual field by IT cortex, even whole visual hemifields (Desimone & Ungerleider, 1989).

Imagine placing an object in front of an observer, say the letter A. The small receptive fields of neurons in V1 might represent the presence of edges and their orientations, and the larger fields midstream might represent more complex junctions of edges, while the largest fields of IT cortex might represent the letters as wholes, or close to it (there is evidence for specialized areas for letter and word processing,

but for the moment we can assume IT cortex). At each stage neurons encode a specific set of features, and the more clearly they signal that those features are present (and not others), the more confidence the system contains about which letter is present. If additional features that are not associated with the letter A are present, that signal will be exposed to increased noise (Tsotsos, 1990).

For example, imagine adding an H to the display as well (as in the crowding example in Fig. 10.1). In V1, the small receptive fields would separately encode the features of each object, and there would initially be no competition among them. But as the features of both objects travel up the ventral stream, at some point features from both objects will enter the same larger receptive field. If the neuron responded highly to horizontal symmetry or the presence of line intersections, then the signal from the A would remain high. But if the neuron responded to vertical symmetry or letter identity, the signal available that an A was present would be decreased due to the competition between the dueling properties. This competition would continue at higher areas of the ventral stream, which in turn feed back to lower areas in an attempt to clarify the conflict toward a winner.

In summary, if a neuron prefers a certain stimulus, and a nonpreferred object is added to its receptive field, that neuron's responses are altered toward the response it would give for the nonpreferred object (Luck, Chelazzi, Hillyard, & Desimone, 1997; Moran & Desimone, 1985; Reynolds & Desimone, 2003). While these neural recordings are only typically possible in monkeys, a similar effect can also be seen in humans using functional magnetic resonance imaging (fMRI). When shown pictures of objects close together on a display, observers showed lower activation in mid and late ventral stream areas (V4, IT cortex) when the pictures were shown simultaneously next to each other, relative to being shown sequentially (Kastner, DeWeerd, Desimone, & Ungerleider, 1998). This result is consistent with the idea that in the simultaneous condition the pictures compete with each other, lowering the overall activation. The effects were smaller when the pictures were moved farther from each other, which should decrease the likelihood that they would fall within the same receptive fields. The effects were also smaller in earlier visual areas (V1), where receptive fields should be too small to allow competition between the pictures.

In this case of competition we assumed that adding more objects added noise, such that information would be lost. But in many other cases adding more objects or features might add activation that does not strongly spatially overlap across the receptive fields of neurons or across the types of features that those neurons respond to (Fujita, Tanaka, Ito, & Cheng, 1992; Komatsu & Ideura, 1993; Treisman & Gelade, 1980). If so, then all or most of the information could still be represented. But it could create a related case of competition for representation, by producing a "binding" problem. If the features of many objects are encoded simultaneously, and multiple objects are present within the same receptive fields, it may not be clear which features should be associated with each object (Luck, Girelli, McDermott, & Ford, 1997; Reynolds & Desimone, 2003; Treisman, 1996).

For example, some visual search data suggest that if a target is designated as an object that contains a set of features (e.g., a gray vertical line) that are also present individually in distractors (e.g., gray horizontals and black verticals), the task is relatively difficult, even though search for any feature alone is easy and immediate (Treisman & Gelade, 1980). This difficulty of these "conjunction searches" may be due to the ambiguity that would arise in determining whether all features were present on any one object (but see McLeod, Driver, & Crisp, 1988; VanRullen, 2009; and Wolfe, 1994, for evidence that many feature combinations that seem like they should produce binding errors can still lead to efficient search). Another case of this binding problem might be seen in the illusory conjunction effect depicted in Figure 10.1. The brief presentation of the display leads to too many features of too many objects being encoded at once, causing confusion about which features belong to which objects. In summary, this type of competition is not for representation alone, but for a type of representation that allows the features of any one object to remain tied together.

How does the visual system resolve these types of competition for representation? One strategy appears to be to reduce competition by suppression of representations of irrelevant information, or heightening representations of relevant information, allowing it to "win" the competition and inhibit competing possible conclusions. Biasing the competition in this way can help the relevant information to outweigh or exclude irrelevant information, and increase the signal from relevant objects (Desimone & Duncan, 1985; Moran & Desimone, 1985; Luck, Girelli, McDermott, & Ford, 1997; Reynolds & Desimone, 1999; Serences & Yantis,

2006). It would be impossible to isolate every bit of relevant information, but the visual system has several tools available that allow selection of types of incoming information that should be *correlated* with the relevant stimuli.

One tool is the selection of areas of the visual field that contain features relevant to the current task, such as certain colors, luminances, spatial frequencies, types or directions of motion, or orientations (see Wolfe & Horowitz, 2004 for a sample list). For example, making a certain color disambiguate relevant information within a display containing harmful irrelevant information can bias activity in the ventral stream toward objects with that color, according to direct neural recordings (Chelazzi et al., 1998), as well as behavioral (Saenz, Buracas, & Boynton, 2003) and electrophysiological (Anderson, Hillyard, & Muller, 2008) studies using human observers. Another tool is the selection of single locations or objects (Mozer & Vecera, 2005; Scholl, 2001) or multiple locations or objects (Awh & Pashler, 2001; Franconeri et al., 2007). Selection can be based either on top-down factors like the observer's current goals (Folk, Remington, & Johnston, 1992) or default bottom-up factors like stimulus salience (Franconeri, Simons, & Junge, 2004; Itti & Koch, 2001).

Selection can bias competition in powerful ways. One set of studies isolated neurons in V4 that produced different responses for preferred and nonpreferred stimuli. As mentioned earlier, when the preferred stimulus was in the neuron's receptive field, adding the nonpreferred stimulus decreased responses. But giving the monkey an incentive to select the location of the preferred stimulus virtually eliminated this decrease (Reynolds, Chelazzi, & Desimone, 1999). Likewise, selecting the nonpreferred stimulus drove the neuron to respond as if it were the only object in the field as well. A set of similar studies additionally showed that when the nonpreferred stimulus was placed outside of the receptive field of the recorded V4 neuron, selection effects were minimal (Luck, Chelazzi et al., 1997; Moran & Desimone, 1985), suggesting that there are few effects of selection when there is no competition to resolve. As another example of the role of selection in resolving competition, in the Kastner et al. (1998) study discussed earlier, simultaneous presentation of pictures caused competition that lowered overall responses in late ventral stream areas. But asking observers to attend to the location of the pictures reduced the suppressive effect of the other pictures.

Later sections of this chapter argue that competition for representation can cause most or all of the effects depicted in Figure 10.1. We start by dividing the types of limits according to a neurophysiologically inspired division between the ventral visual stream (discussed earlier), which is proposed to focus on object identity processing, and the dorsal visual stream, which is proposed to focus on processing of spatial and action-related properties (Ettlinger, 1990; Goodale & Milner, 1992; Mishkin & Ungerleider, 1982), though many argue that this division is not always clear (Cardoso-Leite & Gorea, 2010; Franz, Gegenfurtner, Bulthoff, & Fahle, 2000). The following section applies competition for representation within the ventral stream to the examples on the left side of Figure 10.1, and the section on "Competition for Representation of Selection Locations in the Dorsal Stream" applies competition for representation in the dorsal stream to the examples on the right side of Figure 10.1.

Competition for Representation of Object Identity in the Ventral Stream

For cueing, it is not clear that the competition account could explain the advantage for the cued location, because there is no competing information from other stimuli that would need to be suppressed. However, as discussed in the section on "Why Is Processing Limited?" there is debate over whether cueing effects can be classified as reflecting resource limitations in the first place.

For visual search, there should be competition for representation among targets and distractors, especially when their features are more similar (Duncan & Humphreys, 1989). This should be especially true when objects are placed close together, which should make competition more likely (Motter & Simoni, 2007). Indeed, for conjunction searches where competition should be expected to be high due to binding errors, spacing objects further apart improves performance (Cohen & Ivry, 1991). This competition could be resolved by selecting single objects or handfuls of objects at once, either by their locations or features (see Wolfe, 1994 for review).

Flanker interference and global interference could be described as failures of selection that fail to prevent competition. For the flanker effect, the amount of interference can be reduced by moving the flankers farther from the target, manipulations that should enhance an observer's ability to select only the target for representation. Note that it is also possible that both target and flankers are processed completely and independently through the

ventral stream, and only at the stage of response selection is there competition for control of action (Deutsch & Deutsch, 1963). However, it is difficult for this response selection account to explain effects of spacing between the letters. Global interference could also be due to a failure of selection, due to a default bias toward selecting larger objects or lower spatial frequencies (Navon, 1977; but see Rijpkema, van Aalderen, Schwarzbach, & Verstraten, 2007, for important limitations).

The attentional blink has also been explained in terms of selection that filters incoming information in order to reduce competition among them (DiLollo, Kawahara, Ghorashi, & Enns, 2005; Olivers, van der Stigchel, & Hulleman, 2007).[1] In the example in Figure 10.1, observers should have a selection "setting" for letters. As soon as the first letter ("A") appears, its identity begins to be processed by selecting that letter's location. But as the next character (a digit) appears, it is processed to some degree, which causes the selection "setting" to involuntarily switch over to a setting for "digits," causing the subsequent letter ("E") to be inhibited, and not reported. At first glance, this story sounds a bit post hoc, but it has impressive supporting data. Within the series of the first letter, intervening digit, and second letter, if the intervening digit is changed to a letter, the "blink effect" disappears and the observer can report *all three* letters with the same level of accuracy, consistent with the idea that the digit had reset the observer's selection settings. In contrast, a "resource" version of the blink account would predict that changing the intervening digit to a letter should have made the task more difficult.

For crowding, if objects are placed too closely to one another, such that their representations do not overlap at early ventral areas (allowing processing of basic features), but do compete at late ventral areas (creating ambiguity about which objects are present), the competition account could predict that the local features of objects would be accessible, but the exact identities and their bindings would be ambiguous (Pelli et al., 2004). Adding salient features that allow selection of one object reduces this crowding effect (Chakravarthi & Cavanagh, 2007). Straddling two crowded objects across the visual hemifield boundary also reduces crowding, presumably because high-level receptive fields are usually restricted to a single hemifield, preventing multiple objects from falling into the same receptive field (Chakravarthi & Cavanagh, 2009; Liu, Jiang, Sun, & He, 2009; Torralbo & Beck, 2008).

Having basic features represented at lower levels of the ventral stream, even without proper binding of features to specific objects at higher levels, may still provide some information about a set of crowded objects. The ability to average orientation and size in Figure 10.1 ("Averaging" panel) may reflect an ability to still combine these jumbled features in a useful way. When objects cannot be isolated within a receptive field, this coarser and more global representation may be the only one available (Choo & Franconeri, 2010; Parkes et al., 2001; see Balas, Nakano, & Rosenholtz, 2009 for a related idea). But when isolation of objects is possible, as in the size averaging example in Figure 10.1, broad selection of the space around an entire collection may also make this global representation available in the same way, by activating the features (e.g., size) of all circles at once (Chong & Treisman, 2005).

A similar global jumble may occur for illusory conjunctions. Because the task requires identification of two numbers flanking the shapes, the scope of selection is necessarily broad, encompassing all of the shapes. Simultaneous encoding of all features of all shapes would lead to precise and noncompetitive representations of their features at early stages of the ventral stream, but the broader receptive fields at higher areas would create ambiguity about the relative bindings of the features to each object (Reynolds & Desimone, 1999). Consistent with this idea, moving the objects closer together, which should make them more likely to fall in the same receptive fields, makes reports of illusory conjunctions more frequent (Cohen & Ivry, 1989; Sohn, Liederman, & Reisnity, 1996). Deficits in selection appear to make this illusion occur even outside of briefly flashed displays. For patients with damage to parietal areas known to participate in the control of visual selection, severe damage has led one patient to a state of enduring experiences of illusory conjunctions in everyday life (Robertson, Treisman, Friedman-Hill, & Grabowecky, 1997).

Inattentional blindness seems to be due more to the power of selection than competition per se. In one classic demonstration, observers were asked to count the basketball passes among a team in white shirts, who were interleaved with a team in black shirts, and subsequently missed a gorilla walking through the game (Simons & Chabris, 1999; see Most et al., 2005, for systematic variations using simpler displays). This result is consistent with selection of locations containing white in order to isolate the bounces made only by that team, leading to amplification of anything white in the display,
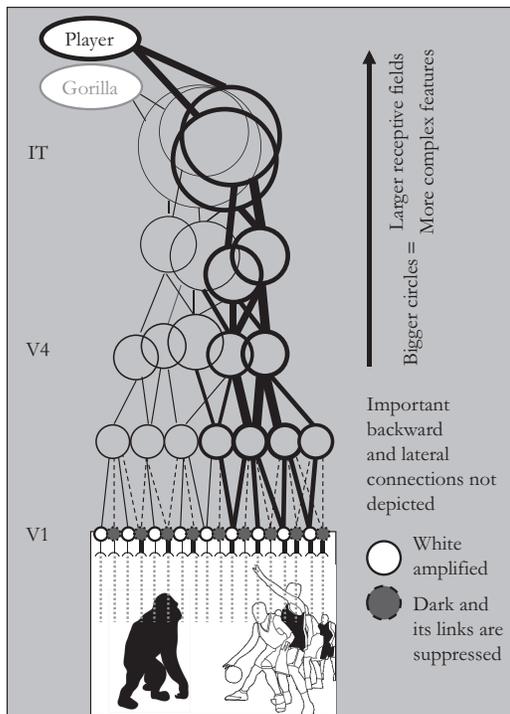
Player

Gorilla

IT

V4

Important
backward
and lateral
connections not
depicted

V1

Bigger circles = Larger receptive fields
More complex features

White
amplified

Dark and
its links are
suppressed

Figure 10.2 An example of how competition for representation could explain missing a gorilla walking through a complex scene. See text for explanation.

relative to suppression of anything black in the display, including the gorilla. Figure 10.2 depicts this process in a cartoon model of the ventral visual system (for clarity, many aspects of the diagram are abstracted, including the scale of receptive fields, the network creating the feature selection, the processing level of competition reduction, and a lack of lateral and backward connections). Two initial sets of feature detectors encode visual information for white and black areas (or contrast polarities) of the scene, and the observer's top-down goal leads to suppression of information from the black detectors (note that the layout of the example scene would allow equally good selection by location, but in the Simons & Chabris 1999 example the gorilla's walking path was interleaved among with the players). This suppression leads the information from the white detectors to win the competition for representation within progressively larger receptive fields. This lack of representation of the gorilla, at least at the highest levels that lead to an explicit memory (Wolfe, 1999), mean that the observer fails to notice the gorilla at all.

Object substitution masking is an important example, because it suggests that competition need

not occur for different objects across space, but can also occur for two objects over time. Object substitution appears to occur when a first wave of visual information containing the target object is overwritten by a second wave containing the mask (see Enns, 2004 for review), such that the mask dots compete for representation with the original objects (and usually win) within a similar set of receptive fields. This competition can be biased toward the original object by cueing the location of the object that will be masked, allowing the selection mechanism to protect the original representation.

Competition for Representation of Selection Locations in the Dorsal Stream

Because these limits on the *right* side of Figure 10.1 do not involve object identification (at least for the aspect of the task that is limited), these limits should not reflect competition for representation within the ventral stream. Instead, they may reflect a similar type of competition in the dorsal stream (Franconeri et al., 2007). Specifically, areas within this stream, notably the lateral intraparietal area (Gottlieb, 2007; Serences & Yantis, 2007) or inferior intraparietal sulcus (Todd & Marois, 2004; Xu & Chun, 2009), may represent currently selected locations in the visual field (Pylyshyn, 1989, 1994). When locations are selected in a complex display, there is a minimum size of a given selection region (Intrilligator & Cavanagh, 2001). That is, in a computer's display, you cannot choose to select a single pixel. Instead, the selection region size (which becomes larger in the periphery) is at minimum roughly one-third the distance of the selected location from fixation (Intrilligator & Cavanagh, 2001). Figure 10.3 depicts sample selection regions for the visual quadrant in panel one (not draw to exact scale), assuming fixation at the center of the figure. In a task requiring selection of multiple relevant "target" locations interleaved with distractor locations, moving locations closer together would cause these selection regions to begin to involuntarily include distractor locations (Franconeri et al., 2007).

To make mattes worse, selection regions are also known to have suppressive surrounds (Bachall & Kowler, 1999; Hopf et al., 2006; Tsotsos, Culhane, Wai, Davis, & Nuflo, 1995), possibly to maximize the amplification of selected information relative to nearby information that might create the most competition within the same receptive fields. As selection regions become closer, these suppressive surrounds would begin to overlap the selection
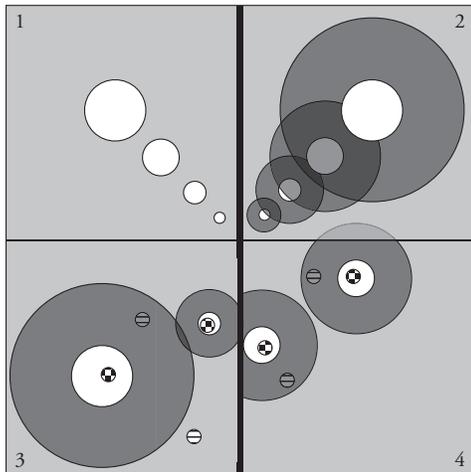
Figure 10.3 Illustration of how competition for representation might explain a limit on the number of locations that can be selected, or the number of objects that can be tracked. See text for explanation.

regions (Franconeri, Jonathan, & Scimeca, 2010; Franconeri, Lin, Pylyshyn, Fisher, & Enns, 2008; Shim, Alvarez, & Jiang, 2008). The second panel of Figure 10.3 simulates this effect. Both of these effects would be worse in the periphery, where selection regions and their associated suppressive surrounds would be larger. The third panel of Figure 10.3 shows how two target locations (checkered circles) could be selected while excluding two distractor locations (striped circles). The fourth panel illustrates how inhibitory surrounds might not cross the vertical visual hemifield boundary, and it might only partially cross the horizontal boundary (Bachall & Kowler, 1999; Cave & Zimmerman, 1997), potentially explaining better performance in some tasks when objects are separated by these boundaries (Alvarez & Cavanagh, 2005; Carlson, Alvarez, & Cavanagh, 2007; Downing & Pinker, 1985). The presence of these boundaries might explain why performance on multiple location selection and multiple-object tracking tasks is limited to about eight objects—it is plausible that within each of the four quadrants, up to two locations or objects can be reliably selected without creating levels of competition that would lead to the loss of an object.

If the limiting factor is competition for representation within a spatial map, we can make a prediction—spacing the objects farther apart should decrease competition and improve performance. Indeed, although both were initially thought to be limited to 4–5 locations or objects (Burkell & Pylyshyn, 1997; Pylyshyn & Storm, 1988), more

recent work shows that these limits can be raised, by keeping the locations or objects as far apart as possible. Moving the objects closer can reduce either limit to 1–2, and moving them farther apart can increase the limit to 8–9 (Alvarez & Franconeri, 2007; Franconeri, Alvarez, & Enns, 2008).

So the limit on selecting multiple locations of moving objects can also be considered competition for representation, this time competing for representation of selection regions with minimum sizes, within the limits of the visual field. But one other effect seems to be inconsistent with this spatially modulated competition as the sole resource in these tasks. When performing a multiple-object tracking task, it is possible to track eight objects, but only when they move very slowly (Alvarez & Franconeri, 2007). When they move faster, performance drops unless the tracking load is reduced, and there are fast speeds at which only one object can be tracked. At first glance it is difficult to explain this limit with competition, because the same number of objects is present in the visual field for slow and fast speeds. There may be another unspecified resource that is drained when objects move faster, requiring a lower number of tracked targets to maintain the same levels of accuracy. But two recent studies suggest that competition over space can explain the speed effects as well. As objects move faster, they cover more distance during a trial (which typically lasts about 5–8 seconds). More distance covered means more opportunity to come too close to other objects, creating competitive interactions that can lead to lower performance. Lowering the number of tracked targets can offset this problem, by lowering the number of these interactions in the first place. This account correctly predicts that increasing the speed of the moving objects does not change performance,[2] as long as it is done in way that does not change the number of object interactions (Franconeri et al., 2008, 2010). These results allow competition for representation to explain the limits in multiple-object tracking tasks.

For rapid counting, the limit on how many objects can be quickly counted (about 4–5) might be similar to the limit on multiple-object selection. The "magic number" might be four instead of eight, because of higher requirements for a precise selection region for the counting process relative to the selection or tracking processes. The requirements of the counting task may require especially clear and noncompetitive selection regions for each object, resulting in a maximum of one object per quadrant. However, there are other plausible

explanations for this fast-counting effect. These explanations include a data-limited cap on our ability to estimate number from correlates of number, such as bumps on a primitive segmentation map, spatial frequency, or display density (Franconeri, Bemis, & Alvarez, 2009; Durgin, 1995). The precision limit of this ability may be enough to tell four from five, but not five from six (Gallistel & Gelman, 1992). Or we might be able to recognize number within small collections according to the stereotypical shapes that they create (1 object is a dot, 2 a line, 3 a triangle, 4 a square or diamond). Larger numbers do not signal prototypical shapes and therefore would not produce efficient performance (Logan & Zbrodoff, 2003; Mandler & Shebo, 1982; Palmeri, 1997; Peterson & Simon, 2000). Because the competition account makes clear predictions about object spacing, there is an experiment that should dissociate these predictions. The competition account predicts that moving the collection of objects into the periphery, or simply one side of the visual field, should increase competition and therefore decrease counting accuracy for small collections.

For tracing, the competition account would state that when a shape becomes sufficiently complex, it is not possible to select multiple areas of that shape with sufficient precision (due to the competition) to differentiate that shape from another interleaved shape. Simplifying the shape, or moving the shape across more of the visual field in ways that decrease competition (e.g., especially across the visual hemifields), should boost performance. Otherwise, there is a salient impression that the task is completed by moving a single mental "spotlight" (which does not incur competition) along the shape or line, and waiting to see where it ends up (Cavanagh, 2004; Jolicoeur et al., 1986; McCormick & Jolicoeur, 1991, 1994; Ullman, 1984; see Houtkamp et al., 2003 for a different explanation).

The last example in Figure 10.1 is spatial relationship perception. This example is tentatively included as a "dorsal" limit, because there is little work exploring the mechanism that allows us to judge the relative locations of a pair of objects. But one strong possibility is that, like in the example of tracing, the position of the locus of selection itself is critical in coding interobject structure. The window of selection might form the "reference frame" for a relationship judgment (Logan, 1995), such that the "left" object is not really on the left of the other object, but on the left of the selection window

encompassing both objects (Biederman, Lescroart, & Hayworth, 2007). This conclusion is not possible unless the selection window spans the two objects being judged, and therefore the selection window itself is a limited visual processing resource. Attempting to place a second window on a second pair of objects would lead to competition about which object is in which window and on which side. A second possibility is that the window of selection only covers one object at a time (Franconeri, Scimeca, Roth, Helseth, & Kahn, 2012). Instead, as selection moves from object to object, the path of the selection window over time is recorded as a vector, allowing the conclusion that the last object viewed is, for example, on the right of whatever was last attended. While these accounts await further empirical investigation, both rely on selection itself as a limited resource.

Conclusions

Competition for representation appears to be a fundamental resource limit for the visual system. It is worth taking a moment to see whether this concept succeeds in unpacking visual "resources" into more concrete terms. Is "competition" just a rephrasing of words like "resource" or "capacity"? Have we merely passed the explanatory buck to the neural level? Competition does explain performance limits across a diverse set of tasks, and more important, it makes surprising and validated predictions about factors that should influence competition, such as object spacing. A generic resource account would have difficulty explaining why identification or multiple location selection improves when densely spaced displays are given more elbow room.

One test for a good account of visual resources mentioned at the start of the chapter is whether one would know what to change about the visual system to increase the resource. For the competition account, it would be more neurons and tighter receptive fields, especially at high-level areas in the ventral stream that typically have the largest fields. So why not do this? Major qualitative changes to a cognitive system (e.g., the advent of recursive processing) might take ages to evolve, but it would seem simple to make quantitative changes by adding more neurons or tightening receptive field sizes. It is possible that this option is limited by physical resource limits like oxygen, glucose, or space in the skull. Too many neurons or connections might be needed to make these changes at the highest level cells in the ventral stream, which are responsible for

representing many different types and combinations of features (Luck & Vecera, 2002). Another likely possibility is that large receptive fields in high-level ventral areas serve an important purpose, reflecting an architecture designed to recognize objects regardless of their size or position on the retina (Lueschow, Miller, & Desimone, 1994). This design should allow the ventral stream to reuse learned patterns across widely varied conditions. Finally, it is possible that increases in visual resources (e.g., tightening receptive field sizes) might not result in processing improvements, because of other limitations on how the visual system's output is used by broader cognitive systems. More detailed information might make it harder to see the forest through trees when learning, notice patterns or correlations, or generate a fast action in response to the world.

Finally, the resource limits of our visual system may cause lower performance in tasks built for psychology labs, but in the natural world several other mechanisms may help us avoid reaching these limits. Even if information is lost through competition, we may still have summary information about the gist or category of a scene (Oliva & Torralba, 2005), the texture or features of objects (Ariely, 2001), or the layout of objects in an environment (Franconeri et al., 2009). When resource limits do become salient, we can guide selection toward information most relevant to our goals (Wolfe & Horowitz, 2004), and selection can be automatically captured by information that is likely to require urgent action (Franconeri & Simons, 2003; Yantis & Hillstron, 1994). Together, such mechanisms may support our perception of an outside world where visual information is fully processed and readily available, without a feeling that we expend resources for the vast majority of everyday visual tasks.

Acknowledgments

Notes

1. This explanation is only part of the larger picture of the resource limitations that underlie the attentional blink effect, which is likely due to a larger set of processing bottlenecks at various stages, including object selection, identification, and memory (see Dux & Marois, 2009, for review).

2. More precisely, increased speed *could* impair performance, just not in a resource-limited way that is relevant for the present argument (e.g., the objects might become blurry). See Franconeri, Jonathan, and Scimeca (2010) for discussion.

References

Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, *16*(8), 637–643.

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track? Evidence for a resource-limited tracking mechanism. *Journal of Vision, 7*(13), 1–10.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162.

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

Bachall, D. O., & Kowler, E. (1999). Attentional interference at small spatial separations. *Vision Research*, *39*, 71–86.

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, *9*(12), 13.1–18.

Biederman, I., Lescroart, M., & Hayworth, K. (2007). Sensitivity to object-centered relations in LOC [Abstract]. *Journal of Vision*, *7*(9), 1030.

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

Burkell, J. A., & Pylyshyn, Z. W. (1997). Searching through subsets: A test of the visual indexing hypothesis. *Spatial Vision*, *11*, 225–258.

Cardoso-Leite, P., & Gorea, A. (2010). On the perceptual/motor dissociation: A review of concepts, theory, experimental paradigms and data interpretations. *Seeing and Perceiving, 23*(2), 89–151.

Carlson, T. A., Alvarez, G. A., & Cavanagh, P. C. (2007). Quadrantic deficit reveals anatomical constraints on selection. *Proceedings of the National Academy of Sciences USA*, *104*(33), 13496–13500.

Carlson, L. A., &Logan, G. D. (2001). Using spatial terms to select an object. *Memory and Cognition*, *29*, 883–892.

Cavanagh, P. (2004). Attention routines and the architecture of selection. In M. Posner (Ed.), *Cognitive neuroscience of attention (pp. 13–28)*. New York: Guilford Press.

Cave, K. R., & Zimmerman, J. M. (1997). Flexibility in spatial attention before and after practice. *Psychological Science*, *8*, 399–403.

Chajut, E., Schupak, A., & Algom, D. (2009). Are spatial and dimensional attention separate? Evidence from Posner, Stroop, and Eriksen tasks. *Memory and Cognition*, *37*(6), 924–934.

Chakravarthi, R., & Cavanagh, P. (2007). Temporal properties of the polarity advantage effect in crowding. *Journal of Vision*, *7*(2),11.1–13

Chakravarthi, R., & Cavanagh, C. (2009). Bilateral field advantage in visual crowding. *Vision Research*, *49*, 1638–1646.

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, *80*, 2918–2940.

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

Choo, H. Y., & Franconeri, S. L. (2010). Visual size averaging of objects unavailable to conscious awareness. *Attention, Perception, and Psychophysics*, *72*(1), 86–99.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation.

Journal of Experimental Psychology: Human Perception and Performance, 21, 109–127.

Cohen, A., & Ivry, R. (1989). Illusory conjunctions inside and outside the focus of attention. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 650–663.

Cohen, A., & Ivry, R. (1991). Density effects in conjunction search: Evidence for a coarse location mechanism of feature integration. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 891–901.

Dehaene, S., & Changeux, J.-P. (2007). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience, 5,* 390–407.

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

Deutsch, J., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, 70, 80–90.

Di Lollo, V., Kawahara, J., Ghorashi, S. M., & Enns, J. T. (2005). The attentional blink: Resource limitation or temporary loss of control? *Psychological Research*, 69, 191–200.

Downing, C. J., & Pinker, S. (1985). The spatial structure of visual attention. In M. Posner & O. Martin (Eds.), *Attention and performance XI* (pp. 171–187). Hillsdale, NJ: Erlbaum.

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

Duncan, J., & Humphreys, G. W . (1989). Visual search and stimulus similarity. *Psychological Review*, 96, 433–58

Durgin, F. H. (1995). Texture density adaptation and the perceived numerosity and density of texture. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 149–169.

Dux, P. E., & Marois, R. (2009). The attentional blink: A review of data and theory. *Attention, Perception and Psychophysics*, 71, 1683–1700.

Enns, J. T. (2004). Object substitution and its relation to other forms of visual masking. *Vision Research*, 44, 1321–1331.

Enns, J. T., & Di Lollo, V. (1997). Object substitution: A new form of visual masking in unattended visual locations. *Psychological Science*, 8(2), 135–139.

Ettlinger, G. (1990). "Object vision" and "spatial vision": The neuropsychological evidence for the distinction. *Cortex, 26*(3), 319–41.

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.

Franconeri, S. L., Alvarez, G. A., & Enns, J. T. (2007). How many locations can you select? *Journal of Experimental Psychology: Human Perception and Performance, 33*(5), 1003–1012.

Franconeri, S. L., Bemis, D. K., & Alvarez, G. A. (2009). Number estimation relies on a set of segmented objects. *Cognition, 113*, 1–13.

Franconeri, S. L., Jonathan, S., & Scimeca, J. M. (2010). Tracking multiple objects is limited only by object spacing, not speed, time, or capacity. *Psychological Science*, 21(7), 920–925.

Franconeri, S. L., Lin, J., Pylyshyn, Z. W., Fisher, B., & Enns, J. T. (2008). Multiple object tracking is limited by crowding, but not speed. *Psychonomic Bulletin and Review*, 15(4), 802–808.

Franconeri, S. L., Scimeca, J. M., Roth, J. C., Helseth, S. A., & Kahn, L. (2011). Flexible visual processing of spatial relationships. *Cognition*, 122, 210–227.

Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception and Psychophysics*, 65(7), 999–1010.

Franconeri, S. L., Simons, D. J., & Junge, J. A. (2004). Searching for stimulus-driven shifts of attention. *Psychonomic Bulletin and Review*, 11(5), 876–881.

Franz, V. H., Gegenfurtner, K. R., Bülthoff, H. H., & Fahle, M. (2000). Grasping visual illusions: No evidence for a dissociation between perception and action. *Psychological Science*, 11(1): 20–25.

Fujita, I., Tanaka, K., Ito, M., & Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature, 360*, 343–346.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*, 43–74.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neuroscience*, 15(1), 20–25.

Gottlieb, J. (2007). From thought to action: The parietal cortex as a bridge between perception, action, and cognition. *Neuron*, 53, 9–16

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Reviews Neuroscience*, 27, 649–677.

Hollingworth, A. (2006). Visual memory for natural scenes: Evidence from change detection and visual search. *Visual Cognition*, 14, 781–807.

Hopf, J. M., Boehler, C. N., Luck, S. J., Tsotsos, J. K., Heinze, H. J., & Schoenfeld, A. M. (2006). Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. *Proceedings of the National Academy of Sciences USA, 103*, 1053–1058.

Houtkamp, R., Sprekreijse, H., & Roelfsma, P. R. (2003). A gradual spread of attention during mental curve tracing. *Perception and Psychophysics*, 65, 1136–1144.

Huang, L., & Pashler, H. (2005). Attention capacity and task difficulty in visual search. *Cognition*, 94, B101–B111.

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

Itti, L., &Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience, 2*(3), 194–203.

Jolicoeur, P., Ullman, S., & Mackay, M. (1986). Curve tracing: A possible basic operation in the perception of spatial relations. *Memory and Cognition*, 14, 129–140.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

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

Komatsu, H., & Ideura, Y. (1993). Relationship between color, shape, and pattern selectivities in the inferior cortex of the monkey. *Journal of Neurophysiology*, 70, 677–694.

Liu, T., Jiang, Y., Sun, X., &He, S. (2009). Reduction of the crowding effect in spatially adjacent but cortically remote visual stimuli. *Current Biology*, 19(2), 127–132.

Logan, G. D. (1994). Spatial attention and the apprehension of spatial relations. *Journal of Experimental Psychology: Human Perception and Performance*, 20(5), 1015–1036.

Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology*, 28(2), 103–174.

Logan, G. D., & Zbrodoff, N. J. (2003). Subitizing and similarity: Toward a pattern-matching theory of enumeration. *Psychonomic Bulletin and Review*, 10, 676–682.

Luck, S. J., Chelazzi, L., Hillyard, S. A., &Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology*, *77*, 24–42.

Luck, S. J.Girelli, M., Mc Dermot, M. T., & Ford, M. A. (1997). Bridging the gap betwen monkey neurophysiology and human perception: An ambiguity resolution theory of visual selective attention. *Cognitive Psychology*, *33*, 64–87.

Luck, S. J., &Thomas S. J. (1999). What variety of attention is automatically captured by peripheral cues? *Perception and Psychophysics*, *61*, 1424–1435.

Luck, S. J., &Vecera, S. P. (2002). Attention. In H. Pashler (Series Ed.) & S. Yantis (Volume Ed.), *Stevens' handbook of experimental psychology, Vol.1. Sensation and perception* (3rd ed., pp. 235–286). New York: Wiley.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Lueschow, A., Miller, E. K., & Desimone, R. (1994). Inferior temporal mechanisms for invariant object recognition. *Cerebral Cortex, 5*, 523–531.

Mack, A., & Rock, I. (1998). *Inattentional blindness*. Cambridge, MA: MIT Press.

Maioli, C., Benaglio, I., Siri, S., Sosta, K., & Cappa, S. (2001). The integration of parallel and serial processing mechanisms in visual search: Evidence from eye movement recording. *European Journal of Neuroscience*, *13*, 364–372.

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General*, *111*, 1–22.

McCormick, P. A., & Jolicoeur, P. (1991). Predicting the shape of distance functions in curve tracing: Evidence for a zoom lens operator. *Memory and Cognition*, *19*, 469–486.

McCormick, P. A., & Jolicoeur, P. (1994). Manipulating the shape of distance effects in visual curve tracing: Further evidence for the zoom lens model. *Canadian Journal of Experimental Psychology*, *48*, 1–24.

McLeod, P., Driver, J., & Crisp, J. (1988). Visual search for a conjunction of movement and form is parallel. *Nature, 332*,154–155.

Mishkin, M., & Ungerleider, L. G . (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioral and Brain Research*, *6*(1), 57–77.

Mitroff, S. R., Simons, D. J., & Levin, D. T. (2004). Nothing compares two views: Change blindness can occur despite preserved access to the changed information. *Perception and Psychophysics*, *66*, 1268–1281.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*, 782–784.

Mordkoff, J. T. (1996). Selective attention and internal constraints: There is more to the flanker effect than biased contingencies. In A. Kramer, M. G. H. Coles, & G. Logan (Eds.), *Converging operations in the study of visual selective attention* (pp. 483–502). Washington, DC: American Psycholgical Association Press.

Most, S. B., Scholl, B. J., Clifford, E., & Simons, D. J. (2005). What you see is what you set: Sustained inattentional blindness and the capture of awareness. *Psychological Review*, *112*, 217–242.

Motter, B. C., & Simoni, D. A. (2007). The roles of cortical image separation and size in active visual search performance. *Journal of Vision*, *7*(2), 6.1–15.

Mozer, M. C., & Vecera, S. P. (2005). Object- and space-based attention. In L. Itti, G. Rees, & J. Tsotsos (Eds.), *Neurobiology of attention* (pp. 130–134). New York: Elsevier.

Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception and Psychophysics*, *70*(7), 772–788.

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

Navon, D. (1984). Resources—A theoretical soup stone? *Psychological Review*, *91*(2), 216–234.

Neisser, U., & Becklen, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology*, *7*, 480–494.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource limited processes. *Cognitive Psychology*, *7*, 44–64.

Oliva, A., &Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, *155*, 23–36

Olivers, N. L., van der Stigchel, S., & Hulleman, J. (2007). Spreading the sparing: Against a limited-capacity account of the attentional blink. *Psychological Research*, *71*, 126–139

Palmer, J. (1994). Set-size effect in visual search: The effect of attention is independent of the stimulus for simpler tasks. *Vision Research*, *34*, 1703–1721.

Palmer, J. (1995). Attention in visual search: Distinguishing four causes of set-size effects. *Current Directions in Psychological Science*, *4*, 118–123.

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 324–354.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*(7), 739–744.

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

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision*, *4*(12), 1136–1169.

Peterson, S. A., & Simon, T. J. (2000). Computational evidence for the subitizing phenomenon as an emergent property of the human cognitive architecture. *Cognitive Science*, *24*(1), 93–122.

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

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

Pylyshyn, Z. (1994). Some primitive mechanisms of spatial attention. *Cognition, 50*, 363–384.

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

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 849–860.

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience*, *19*, 1736–1753.

Reynolds, J. H., &Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron*, *24*, 19–29.

Reynolds, J. H., &Desimone, R. (2003). Interacting roles of attention and visual salience in V4. *Neuron*, *37*(5), 853–863.

Rijpkema, M., van Aalderen, S., Schwarzbach, J., & Verstraten, F. A. J. (2007). Beyond the forest and the trees: Local and global interference in hierarchical visual stimuli containing three levels. *Perception, 36*(8), 1115–1122.

Robertson, L., Treisman, A., Friedman-Hill, S., & Grabowecky, M. F. (1997). The interaction of spatial and object pathways: Evidence from Balint's syndrome. *Journal of Cognitive Neuroscience, 9*, 295–317.

Saenz, M., Buracas, G. T., & Boynton, G. M . (2003). Global feature-based attention for motion and color. *Vision Research*, *43*, 629–637.

Scott-Brown, K., Baker, M. J., & Orbach, H. (2000). Comparison blindness. *Visual Cognition, 7*, 253–267.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*(1/2), 1–46.

Serences, J. T., &Yantis, S. (2006). Selective visual attention and perceptual coherence. *Trends in Cognitive Sciences*, *10*, 38–45.

Serences, J. T., &Yantis, S. (2007). Representation of attentional priority in human occipital, parietal, and frontal cortex. *Cerebral Cortex*, *17*, 284–293.

Shim, W. M., Alvarez, G. A., & Jiang, Y. V. (2008). Spatial separation between targets constrains maintenance of attention on multiple objects. *Psychonomic Bulletin and Review*, *15*(2), 390–397.

Shiu, L-P., & Pashler, H. (1994). Negligible effect of spatial precuing in identification of single digits.*Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1037–1054.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*, 1059–1074.

Sohn, Y-S., Liederman, J., & Reinitz, M. T. (1996). Division of inputs between hemispheres eliminates illusory conjunctions. *Neuropsychologia*, *34*, 1057–1068.

Todd, J., & Marois, R. (2004). Capacity limit of visual sort-term memory in human posterior parietal cortex. *Nature*, *428*(15), 751–754.

Torralbo, A., &Beck, D. M. (2008). Perceptual-load-Induced selection as a result of local competitive interactions in visual cortex. *Psychological Science*, *19*(10), 1045–1050.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*, 171–178.

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

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*, 107–141.

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited capacity preattentive stage in vision. *Psychological Review*, *101*, 1–23.

Tsotsos, J. K. (1990). Analyzing vision at the complexity level. *Behavioral and Brain Sciences*, *13*, 423–469.

Tsotsos, J. K., Culhane, S. W., Wai, Y. L., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence*, *78*, 507–547.

Ullman, S. (1984). Visual routines. *Cognition*, *18*(1–3), 97–159.

VanRullen, R. (2009). Binding hardwired vs. on-demand feature conjunctions. *Visual Cognition*, *17*(1–2), 103–119.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, *1*(2), 202–238.

Wolfe, J. M. (1998). What can 1,000,000 trials tell us about visual search? *Psychological Science*, *9*(1), 33–39.

Wolfe, J. M. (1999). Inattentional amnesia. In V. Coltheart (Ed.), *Fleeting memories: Cognition of brief visual stimuli* (pp. 71–94). Cambridge, MA: MIT Press.

Wolfe, J. M. (2007). Guided search 4.0: Current progress with a model of visual search. In W. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–119). New York: Oxford University Press.

Wolfe, J., Alvarez, G., Rosenholtz, R., Oliva, A., Torralba, A., Kuzmova, Y., & Sherman, A. M. (2008). Search for arbitrary objects in natural scenes is remarkably efficient. *Journal of Vision*, *8*(6), 1103–1103.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*, 1–7.

Xu, Y., &Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends in Cognitive Sciences*, *13*, 167–174

Yantis, S., & Hillstrom, A. P. (1994). Stimulus-driven attentional capture: Evidence from equiluminant visual objects. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 95–107.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233–235.