In *Phenomenal Qualities: Sense, Perception, and Consciousness*, P. Coates and S. Coleman, eds. Oxford: Oxford University Press, 2015. pp. 347-375. Final version.

A Function-Centered Taxonomy of Visual Attention

Ronald A. Rensink

Departments of Psychology and Computer Science

University of British Columbia

Vancouver, Canada

Correspondence concerning this paper may be addressed to R.A. Rensink, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver BC V6T 1Z4, Canada. Email: rensink@psych.ubc.ca or rensink@cs.ubc.ca.

Abstract

It is suggested that the relationship between visual attention and conscious visual experience can be simplified by distinguishing different aspects of both visual attention and visual experience. A set of principles is first proposed for any possible taxonomy of the processes involved in visual attention. A particular taxonomy is then put forward that describes five such processes, each with a distinct function and characteristic mode of operation. Based on these, three separate kinds—or possibly grades—of conscious visual experience can be distinguished, each associated with a particular combination of attentional processes.

Key words

attention; binding; change blindness; coherence; consciousness; inattentional blindness; taxonomy; vision; visual experience; visual perception

A Function-Centered Taxonomy of Visual Attention

It is often said that appearances can be deceiving. This is certainly true in regards to what they convey about the world. But appearances can also deceive in regards to what they convey about the processes that create them. For example, our visual experience of the world is so immediate and effortless that it tends to engender a belief that scene perception results from a single unitary system that lets us immediately experience everything in our field of view. But a host of experiments in vision science have shown this not to be the case. Instead, scene perception appears to rely on the co-ordinated operation of several systems: an *early* system that rapidly creates a dense and volatile[1] representation across much of the visual field, an *attentional* system that selects a small part of this and forms it into a coherent visual object, and a *setting* system that guides this selection so that the right item is attended at the right time (e.g., Rensink, 2000a, 2010).

The possibility considered here is that this is not the end of the line—that we may likewise have an incorrect belief about visual attention itself. Although "attention" is easy enough to understand at a subjective level, it has been notoriously difficult to characterize in an objective way (e.g., Allport, 1993; Chun, Golomb, & Turk-Browne, 2011; Jennings, 2012). Difficulties have also been encountered in reconciling various proposals about how attention relates to conscious perceptual experience (cf. e.g., Cohen, Cavanagh, Chun, & Nakayama, 2012; De Brigard & Prinz, 2010; Kentridge, 2011; Koch & Tsuchiya, 2007; Lamme, 2003). It has been suggested that many of these difficulties exist because—contrary to common belief—"attention" does not refer to a single process, but is an umbrella term referring to *several* processes (Allport, 1993; Treisman, 1969). This paper explores one way of developing this idea, and discusses how it might help us better understand conscious visual experience.

[1] "Dense" means that when a quantity is present in some area, it exists at most points in that area. "Volatile" means that the representation is not robust, being overwritten by subsequent input, or in the absence of that, quickly dissipating. Dissipation is typically complete within a few hundred milliseconds. The trace—while it remains—corresponds with iconic memory (Rensink, 2002).

In particular, it is suggested here that visual attention may best be viewed in terms of the co-ordinated operation of a set of processes, each with a distinct function and underlying mechanism. A key concern is then to develop a taxonomy that can describe each process and relate it to others. Towards this end, a set of constraints is first proposed on the nature of any such taxonomy, centered around the function of each process—viz., the kind of structure it outputs. A particular set of processes is then presented that is consistent with these constraints and that provides a coherent grouping of many (if not most) experimental results to date. Finally, it is suggested that a similar fractionation may apply to conscious visual experience. Among other things, this creates the possibility of reducing the problematic relationship between visual attention and visual experience to a set of simpler issues, each concerned with the relationship between a particular kind of attentional process and a particular kind of visual experience.

1. The Nature of Visual Attention

Performance in many visual tasks is governed by a factor within the observer that enables certain operations to be carried out, but is limited in some way. For example, when keeping track of several automobiles in traffic, only a small number can be handled simultaneously; if more are attempted, performance begins to fail. This limited factor is generally referred to as *attention*. But what exactly is it?

Considerable work has been carried out on this question over the years (see e.g., Itti, Rees, & Tsotsos, 2005; Jennings, 2012; Pashler, 1999; Wright, 1998). At various times, visual attention has been associated with things such as clarity of perception, consciousness, or a limited "resource" that enables particular kinds of operations to be carried out (see Hatfield, 1998). But perhaps the greatest amount of progress has been achieved by focusing on the idea of *selection* (Broadbent, 1982).

In what follows, an attentional process is taken to be one that is *contingently selective, controlled on the basis of global considerations* (Rensink, 2013)—e.g., tracking a particular item based on its estimated importance.[2] Global considerations include not only cognitive factors such as the importance for the task at hand, but also perceptual factors such as *salience*—the visual distinctiveness of an item with respect to all others in the visual field. These are largely handled via two kinds of attentional control: *endogenous* (via top-down cognitive mechanisms subject to conscious volition), and *exogenous* (via bottom-up perceptual mechanisms that operate automatically, although still in terms of global considerations).

In this view, "attention" is more an adjective than a noun—an attentional process is one that is selective and subject to a particular kind of control; "paying attention" is exerting that control, resulting in a particular kind of selection. One advantage of this characterization is that it allows "attention" to be implemented in different ways—there need not be a single process that can be identified with it, nor a single site where it operates (cf. Allport, 1993; Tsotsos, 2011, ch. 1). This characterization excludes processes such as the transduction of light by photoreceptors: although this process is selective (in that it has a differential sensitivity to wavelength), its selectivity is not contingent. On the other hand, this characterization includes any globally-controlled process of limited capacity (such as storing information into visual short-term memory), since limited capacity necessarily results in selectivity of some kind. And although this definition can apply to the controlled allocation of a single resource or process, it is not limited to this—it can also apply to the control of several processes, provided this is done in a co-ordinated way.

This characterization departs somewhat from the traditional notion of attention as a simple "stuff" or "resource". This notion, however, does not always help make sense of experimental results, nor does it always lead to interesting experimental questions (Allport, 1989; Franconeri, 2013; Navon, 1984). Even more importantly, the traditional view has been unable to engender a comprehensive framework capable of providing a good understanding of attention (see e.g.,

---

[2] This is somewhat similar to the proposal of Jennings (2012), except that no explicit mention is made here of a "subject" that controls the process. The emphasis here is instead on objective factors that are nonlocal, such a responding to the brightest dot in an image.

Allport, 1993) or its connection to conscious experience. It may therefore be time to give serious consideration to the alternative.

2. Constraints on Potential Taxonomies

If different kinds of attentional process exist, a taxonomy of some kind could help delineate what these are and how they relate to each other. But creating such a system involves addressing several taxonomic issues, such as the nature of the main categories, and the appropriate granularity for each. How might this be done?

2.0. General Principles

A reasonable constraint for any potential taxonomy is that it be based on principles that are relatively general, and so unlikely to significantly change as new facts are discovered. Several such principles appear relevant:

1. Function-centeredness: The specification of each process must center around its *function*—e.g., selecting information from a particular part of the visual field, or explicitly linking estimates of orientation and color at a given location. Such a focus provides greater generality than if the taxonomy were organized around type of perceptual effect or experimental task used. As mentioned earlier, the function must be more than just selective: it must also be controllable on the basis of global considerations, such as those conveyed by task instructions.

2. Maximal functionality: The processes contained in the taxonomy must include—in one form or other—as many useful distinctions as possible in terms of function. This might include, for example, selection of items in a given area of space, or of a particular color. (Assuming that these can be controlled via global considerations.) No important ability should be neglected. Following the principle of function-centeredness, the articulation of these should be done in terms of function alone, without regard to implementation.

3. Minimal mechanism: The taxonomy must posit as few underlying mechanisms as possible. (In essence, this is a form of Occam's razor.) Here, *mechanism* refers to the set of operations, implemented in a neural substrate, that carry out the function under consideration. When description is in terms of observable properties alone (e.g., time taken, number of operands), this can be viewed as a *specification*; a more complete characterization would include the *algorithm* used, along with a description of the representation(s) involved. Some mechanisms can support several functions, although not necessarily concurrently (e.g., a computer operating system). The complete set of functions associated with a mechanism—along with the mechanism itself—corresponds to a single attentional *process*.

4. Computational analysis: Each process should be described in terms of (i) function, (ii) perceptual effects, (iii) mechanism (algorithm and representation), and (iv) neural implementation. The need for a functional description is clear from the above. The requirement for perceptual effects operationalizes these functions so as to guarantee they can be measured. The other two requirements are part of the computational analysis of any visual process (Marr, 1982); the result could be viewed as an *augmented computational theory* of attention. A complete analysis from all four perspectives is more of an ideal than a reality—in practice, only some of these may be possible. But following the principle of function-centeredness, analysis at the functional level should always be included.

2.1. Distinctions Concerning Function

The maximal-functionality principle requires that a taxonomy make as many functional distinctions as possible. But what should these be? And how should they be grouped? One possibility (Chun et al., 2011) is to begin by distinguishing between functions involving information *external* to the observer (e.g., collecting it from a selected location in the environment) and information that is *internal* (e.g., transmitting it along particular pathways). In what follows, consideration will be limited to external functions, which are better understood and arguably more closely related to perceptual experience.

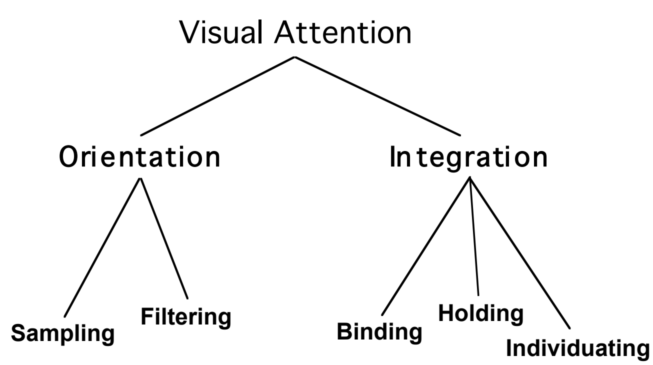
External functions can be subdivided (Rensink, 2009, 2013) into *orientation* (selective access to a particular set of data from the environment) and *integration* (selection of a particular set of spatiotemporal associations for this data). Orientation can in turn be subdivided into two sub-functions:

- *sampling* (collection of sufficient information from the world)
- *filtering* (discarding of unnecessary information).

Ideally, these result in the collection of exactly the information necessary for the task at hand. For integration, various sub-functions can likewise be defined:

- *holding* (association across both time and space)
- *binding* (association across space alone; minimal temporal extent)
- *individuating* (association across time alone; minimal spatial extent)

Figure 1 shows the result[3]. Nothing prevents these functions from being subdivided further; indeed, the principle of maximal functionality would require making any distinction that is useful. For example, binding could be separated into more specialized forms such as linking items with a given color, or associating an item to a particular location in space. The ultimate granularity of such distinctions will likely be determined by practical considerations. In any event, the relationships established between these functions on the basis of such refinement form a "shadow taxonomy", which can specify the relationships of any processes subsequently posited as being distinct.



[3] In the nomenclature used here, individual processes end in "-ing", the participle indicating their status as active entities; higher-level, more abstract groupings end in "-tion".

Figure 1. Potential subdivisions of external functions associated with visual attention. This reflects functional aspects only; these divisions correspond to different processes only if the mechanisms that underlie them are different. Many of these distinctions could be further subdivided.

2.2. Distinctions Concerning Mechanism

The minimal-mechanism principle requires assuming as few underlying mechanisms as possible. This may be done by initially assuming a single underlying mechanism and positing separate mechanisms only when there is sufficient evidence for doing so. Two general kinds of technique are helpful here. The first is based on *dissociation*—a manipulation (typically in the input) that affects one process but not the other. Two processes are considered to differ (or more precisely, involve mechanisms not used by the other) if they show a *double dissociation*—a pair of dissociations such that each process is affected separately (see e.g., Chun, 1997). For example, if each process operates on an entirely different kind of input (one on sound, one on light, say), it will be affected only by a change in that input.

The second kind of technique is based on *dual-task interference* (see e.g., Braun, 1994; VanRullen, Reddy, & Koch, 2004). Here, performance on two tasks carried out concurrently is compared to when they are carried out separately one at a time. If no detriment exists, the processes involved do not use a common resource, and thus do not draw upon common mechanisms. A popular form of this is the *attentional operating characteristic,* (Sperling & Dosher, 1986), which describes the extent to which two attentional tasks interfere with each other.

3. A Provisional Taxonomy

Although the considerations above constrain any potential taxonomy, they are not enough to specify it uniquely. A provisional candidate is therefore suggested here, consistent with all the above constraints and capable of organizing most known perceptual effects (including deficits) related to visual attention (Rensink, 2013). Although unlikely to completely capture all aspects of attention, it will serve to keep discussion focused, and provide a basis of comparison for any proposed improvement.

Note that some of the processes described have a nonattentional component to their control—for instance, they could be controlled *reflexively*, via a precompiled routine that operates automatically on the basis of local visual properties such as edges of sufficient contrast[4]. This does not prevent them from being attentional, however, in that they can still be controlled on the basis of global considerations whenever necessary.

3.0. Attentional Sampling

A. Function

Visual perception begins with the pickup of information via the selective *sampling* of incoming light by the eye. Because the eye has high resolution only in the few degrees around the point of *fixation*, it must be continually repositioned via brief jumps, or *saccades*; when controlled appropriately, these allow the right information to be obtained from the environment at the right time (Ballard, Hayhoe, Pook, & Rao, 1997; Carpenter, 1988). Although the control of this process can be reflexive, it often involves global considerations—e.g., fixating or pursuing the object needed for the task at hand. In such situations, this process is attentional in the sense used here.[5] It is also attentional in the traditional sense, which refers to it as *overt attention*. To make more explicit its functional role, it is referred to here as *attentional sampling*.

Other selective processes exist that are entirely internal to the observer; their operation is often referred to collectively *as covert attention*. Overt and covert systems are only partly correlated: they need not—and often do not—act on the same information (e.g., Juan, Shorter-Jacobi, & Schall, 2004). Put another way: if an observer is fixating (and thus sampling) a given

---

[4] Reflexive control is similar to exogenous control (Section 1) in that both are automatic, and can be driven by the contents of the image. However, exogenous control involves *global* considerations (e.g., processing the highest-contrast item in an image as part of a given task) whereas reflexive control only involves *local* ones (e.g., processing any item with sufficient contrast, no matter what). Reflexive control would likely be overridden by global control in most situations. But it would generally be difficult to empirically determine whether the control used in any particular situation is exogenous or reflexive.

[5] It may be worth emphasizing that according to the definition used here, a selective process is attentional only when global control is in effect.

item on the basis of some global consideration, it does not necessarily follow that they are "attending" to it in all other possible ways.

B. Perceptual Effects

Attentional sampling has been the subject of considerable work by vision scientists, in large part because its external (overt) character allows direct observation and manipulation of its operation. Selectivity can be measured when eye movements are prevented, such as having the observer view an image during a brief flash. Acuity and color perception are best in the central few degrees (or *foveal* area), falling off rapidly with eccentricity in the outer parts (or *periphery*). Information is thus of maximal resolution. Conversely, motion perception is poor in the fovea and better in the periphery (see e.g., Barlow & Mollon, 1982).

C. Mechanism

The systems underlying sampling are reasonably well understood. Incoming light is picked up by two interleaved arrays of photoreceptors: *rods* and *cones*. Rods are located mostly in the periphery, cones—which alone can distinguish color—in the fovea. The eye itself is moved via three pairs of muscles, controlled via areas in the brain such as the superior colliculus and the frontal eye fields (see e.g., Crowne, 1983); these are in turn controlled via several different neural pathways. (For more information, see e.g., Henderson & Hollingworth, 1998; Krauzlis, 2005).

3.1. Attentional Filtering (Gating)

A. Function

Although a vast amount of information is picked up via sampling, most is irrelevant for any given task. This irrelevant information can significantly degrade performance, essentially acting as a form of noise (see e.g., Rensink, 2013). As such, it must be filtered out as much as possible, improving the *quality* of the information used. (This is sometimes referred to as applying "selective attention".) A simple way of doing so is *gating*—transmitting only the information at a particular location or containing a particular property, such as color or size. Various subdivisions are possible, depending on the parameter(s) controlled. These include:

- *spatial filtering* (selection of information from a particular region of space; sometimes referred to as "spatial attention")
- *feature filtering* (selection of information containing a simple property, or *feature*; sometimes referred to as "feature-based attention")
- *ocular filtering* (selection of information from a particular eye)

An important issue for most of these is the extent of selection. At one extreme, selection can be *diffuse*, with a relatively broad range of inputs. In the case of space, for example, information can be accessed from a large area of the visual field, allowing some processes to speed up by operating in parallel; the downside, however, is that more noise is introduced. Selection can also be *focused*, with input taken from a relatively restricted range—e.g., from a small region of space. This reduces noise, but may cause processing to slow down. (These extremes are sometimes referred to as respectively as *diffuse* and *focused attention*.) The extent of selection essentially involves a trade-off between speed and accuracy, with the optimal choice depending upon details of the task and the environment.

B. Perceptual Effects

Attentional filtering lies at the heart of various effects associated with the *quality* of the information transmitted. These include *interocular suppression* (items not seen if originating in the unattended eye), *enhancement* (attended items have greater apparent contrast), and *inattentional blindness* (items not seen if not attended). All can be characterized in an objective way. (For details on these and related effects, see e.g., Itti et al. (2005) and Rensink (2013).)

Several deficits can be traced to difficulties with filtering. Damage to the right posterior parietal cortex can result in *neglect*—the absence of visual experience in some part of space (Bartolomeo & Chokron, 2002; Bisiach, 1993). A related condition is *extinction*, where an item vanishes when a competing item is placed on the opposite side of the visual field. Both deficits appear to result from a failure to gate information from the appropriate area (or representation of the item), causing either an outright failure to access the information it contains, or at least a slowdown of processing (Posner et al., 1984).

C. Mechanism

Although the phenomena above superficially have little in common, several commonalities exist in their mode of operation:

- can be controlled on the basis of global considerations
- inputs can be switched very quickly (typically, within 30-50 milliseconds)
- selection is of simple properties (e.g., spatial locations or simple features)
- for space, a contiguous area is involved, akin to a spotlight (1° in size at fixation; increases with retinal eccentricity)
- transmitted information is poorly localized (can't precisely establish position on retina)

Although the parameters that *control* filtering are simple, the structures[6] *acted upon* are not, generally having at least some degree of organization (see e.g., Driver, Davis, Russell, Turatto, & Freeman, 2001; Rensink, 2013). For example, filtering can be affected by background structures (*segments*) formed on the basis of luminance or texture boundaries at early "preattentive" stages of vision, before attentional filtering has had time to act (Driver et al., 2001). Transmitted information can be measurements of relatively complex structures of limited extent (*proto-objects*) formed at early levels (Rensink & Enns, 1995). Consistent with this, these effects tend to operate over perceived rather than "raw" retinal space (Robertson & Kim, 1999).

3.2. Attentional Binding

A. Function

Ideally, the representations of all properties in an image that are relevant for a task should be explicitly associated or linked so as to adequately capture the structure of the world at each

[6] As used here, "structure" can refer either to structure in the external world or the corresponding structure that is part of representational content. It will hopefully be clear from context which meaning is intended.

moment—i.e., they should be *bound*. Some degree of reflexive binding is already apparent in the creation of preattentive background segments and proto-objects (Section 3.1). But to better capture structure, more sophisticated control is often needed. For example, if a line segment can be assigned equally well to either of two groups on the basis of purely local factors, determination of its placement will need to be based on non-local considerations. In other words, it will need to involve *attentional binding*.

As in the case of filtering, several kinds of binding can be distinguished, many of which can be subdivided in turn. For example:

- *feature binding*
  - across-feature binding (e.g. color and orientation)
  - within-feature binding (e.g. horizontal and vertical lines in a "T")
- *part binding* (or *connecting*)
  - binding across space (i.e., grouping)
  - binding across levels of hierarchical structure
- *position binding* (or *positioning*)
  - binding to a precise location in space

Attentional binding is often associated with filtering. For example, the feature integration theory of Treisman & Gelade (1980) posits that the linking of different kinds of features in an item (e.g., blue and horizontal) occurs via the focused gating of information over a small area at their location, which enables their representations to be activated simultaneously. But from a functional point of view, binding differs substantially from filtering: it is concerned not with *access*, but with *construction.*

B. Perceptual Effects

As in the case of filtering, effects involving attentional binding show up in various ways, all of which can be characterized objectively. But rather than involving the *quality* of information, they involve—via success or failure—its *integration*. These include *conjunction detection* (detection of items with a unique combination of features), *illusory conjunctions* (inappropriate linking of features in briefly-presented items), and *repetition blindness* (failure to distinguish similar stimuli within a brief time). (For more information on these and related effects, see e.g., Kanwisher, Yin, & Wojciulik (1999), Rensink (2013), and Wright (1998).)

Several perceptual deficits can be ascribed to failures of binding. For example, damage to the inferior temporal lobe can result in *integrative agnosia*, an inability to perceive overall shape or configuration; perception is only of simple features, such as color or texture (Farah, 2004; Riddoch & Humphreys, 1987).

C. Mechanism

Attentional binding creates representations that support processes such as the recognition of complex shapes and characters. These appear to depend on a mechanism—or set of mechanisms—having the following characteristics:

- can be controlled on the basis of global considerations
- medium-speed operation (complete in about 100-150 milliseconds)
- operates on organized structures (e.g., segments, proto-objects)
- involves only a small number of such structures at any time

In some cases attentional binding may be carried out via filtering (e.g., Treisman & Gelade, 1980). In other cases, however—such as those involving nonlocal structure—different mechanisms are used (e.g., Maddox, Ashby, & Waldron, 2002; VanRullen et al., 2004).

3.3. Attentional Holding (Stabilizing)

A. Function

When a physical object changes over time (e.g., a bird taking flight), it is often useful to represent it not as a set of unrelated structures, but as a single persisting object. Continuity of this sort can be captured via a representation that is *coherent* (or *stabilized*). Here, properties are linked not only across space but also across time, so as to refer to an object with a substrate that endures, even across eye movements (Kahneman, Treisman, & Gibbs, 1992; Rensink, 2000a)[7]. Owing to the complexity of the processes involved in constructing the underlying representation, such *holding* may require a considerable amount of computational resources. It is therefore unlikely that much of it is done reflexively.

Two kinds of continuity can be distinguished: *perceptual* continuity, where the perceptual representation of an object is continually maintained, and *conceptual* continuity, where this representation is recreated and matched with a trace in long-term (semantic) memory. The first corresponds to the perception of *dynamic change*, where an object is seen to dynamically transform; the second to *completed change*, where the object is simply perceived as having changed at some point in the past (Rensink, 2002).

B. Perceptual Effects

Attentional holding links properties across both space and time. Effects involving its failure or success include *object-specific preview benefit* (faster detection of stimuli located inside objects in which they appeared previously), *attentional blink* (failure to create a second visual object if presented less than 250-300 milliseconds after the first), and *change blindness* (failure to detect clearly-visible change in an object over time). All can be characterized in an objective way. (See e.g., Jensen, Yao, Street, & Simons (2011) and Rensink (2013) for more complete descriptions of these and related effects.)

Several deficits caused by damage to the occipital region of the brain may be linked to mechanisms that enable holding. Among the more striking of these is *dorsal simultanagnosia*. Here, observers cannot recognize more than one object (or part of one object) at a time, with the rest of the input simply not being experienced (Coslett & Saffran, 1991; Farah, 2004). A variant

---

[7] Eye fixations require an exposure of no more than about 150 milliseconds if eye movements are to be optimally guided (Rayner, Smith, Malcolm, & Henderson, 2009). This suggests that the bound structure generated at each fixation suffices for most aspects of perception. Persistence across eye movements would then involve the representation of a different kind of structure.

is *ventral simultanagnosia*, in which observers cannot recognize more than one object, but can still experience several simple (bound) shapes simultaneously (Farah, 2004).

C. Mechanism

Each of the effects associated with attentional holding appears to involve a mechanism with most (if not all) of the following properties:

- can be controlled on the basis of global considerations
- relatively slow operation (250-300 milliseconds)
- operates on organized structures (e.g., segments, proto-objects)
- accesses up to 3-4 such structures (only a small amount of information from each)
- only one structure in play at a time (accessed structures act as "parts")
- stability across interruptions for several seconds (access to visual short-term memory)

Such a mechanism can represent at most only a few aspects of a physical object at any time. Although the situation here is not as well-understood as it is for the others, indications are that attentional holding is also a distinct process (Rensink, 2013).

One model of this is the *object file,* a temporary representation of (bound) properties that captures the continuity of an object as it transforms or changes position (Kahneman et al., 1992). Another is the *coherence field,* in which information is "held" in a reverberating circuit created by feedforward and feedback connections between selected proto-objects and a higher-level collection point, or *nexus* (Rensink, 2000a, 2001).

3.4. Attentional Individuating (Indexing)

A. Function

It is sometimes useful to *individuate* a physical object—to see it not just as *an* object, but as a *particular* object. This can be important when more than one object must be dealt with, such as determining a “between” relation, or ensuring that the items in an image are processed in an efficient sequence (Pylyshyn, 2003; Ullman, 1984). In contrast to other attentional functions, individuating—or "indexing" (Pylyshyn, 2003)—is not concerned with visual structure per se, but with *process*, e.g., applying a particular operation to a particular item at a particular time.

Although individuating could in principle be done via coherent representation, it is difficult to create more than one such representation at a time (Section 3.3). Fortunately, complete coherence is rarely needed: it is often sufficient to consider an item considered distinct at some point, and then *track* it on the basis of its position[8]. If this can be done successfully, the item can be immediately accessed or operated upon whenever needed.

B. Perceptual Effects

Several effects involve items which persist over time, but for which physical properties are unimportant, or even irrelevant. These include *multiple-object tracking* (immediate report about selected items that move about)*, prioritization of search* (faster search for items in locations shown ahead of time), and *subitizing* (rapid counting of a small number of items). (For further details on these and other related effects, see e.g., Rensink (2013) and Scholl (2009).)

C. Mechanism

The mechanism that underlies the various effects involving individuating appears to have many—if not all—of the following characteristics:

- can be controlled on the basis of global considerations
- operates quickly (approximately 30-50 milliseconds per item)
- operates on organized structures (e.g., segments, proto-objects)
- accesses up to 7-8 such structures (via location of their centers of mass)
- only one overall structure at a time (locations organized into a "virtual polygon")
- based on an environment-centered frame (not "raw" retinal space)

---

[8] Tracking may be a distinct process concerned with temporal continuity only (cf. the separation of binding into several possible processes.) However, invoking the principle of minimal mechanism, tracking will be assumed to be an aspect of individuating until evidence is found for a distinct underlying mechanism.

Coherent representations are not necessarily individuated, suggesting that individuating and holding involve different mechanisms (see e.g., Bahrami, 2003; Scholl, 2009). However, individuating does appear to facilitate the control of filtering, suggesting at least some interaction of these (see Rensink, 2013).

3.5. Nonattentional Processing

As the above survey indicates, all attentional processes appear to be influenced by organized structure determined rapidly at early levels, before any process has had much time to act. Although such effects have sometimes been ascribed to "object-based attention", it may be that they originate at a stage used by *all* attentional processes (or at least, all those beyond sampling). In this view, considerable processing takes place reflexively and rapidly across the visual field, resulting in a substrate upon which attentional processes can draw. An interesting issue is the extent to which an attentional process could affect this substrate itself. In the extreme, it might be able to do so substantially, with the resultant structure possibly reverting to its original form after attentional processing ends (see Rensink, 2009, 2010, 2013).

3.6. Dependencies

How might the various attentional processes depend on each other? The simplest set of relations—those among the processes themselves, whether or not global control is in effect—is shown in Figure 2. Here, sampling and filtering are posited as acting prior to the others, since no process can proceed without adequate input. Reflexive gating of information is possible: for example, everything around the edges in an image could simply be transmitted. But global control would help ensure that the transmitted information is exactly what is needed.

The outputs of sampling and filtering are similar in that both involve a dense array of simple quantities. However, the information transmitted by filtering is not tightly bound to retinal position (Section 3.1). Note that the poor localization of filtered output is a natural consequence of positional invariance, which allows processes such as recognition to give much the same result regardless of the exact position of the target in the image (Tsotsos, 2011).

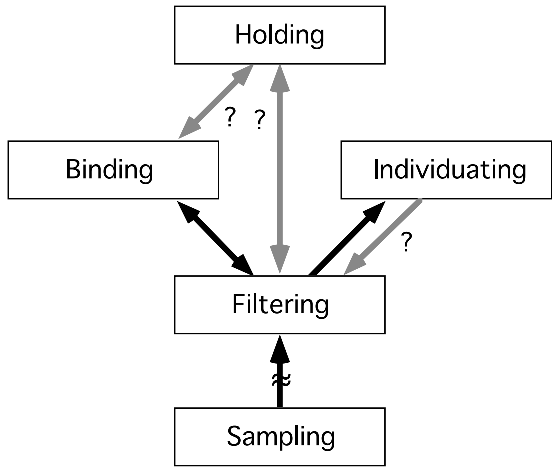


Figure 2. Possible set of dependencies between attentional processes. (Arrows with question marks denote connections for which insufficient evidence exists.) All processes depend on transmission of information at the filtering stage, which in turn depends on sampling. Between sampling and filtering exist processes that enable positional invariance. (Wavy lines indicate that correspondence to retinal position is via a transform to more object-centered coordinates.) It is unclear whether the dependence of holding on filtering is direct, via binding, or both.

The relationships between the other three processes are largely unknown; functional specifications do not completely constrain the situation. For example, although binding involves spatial structure, and holding involves both spatial and temporal structure, the type and extent of spatial structure in each need not be the same. If binding involves more extensive spatial structure than holding, say, its outputs are unlikely to be directly drawn upon by holding. However, many of these dependencies are likely reciprocal—for example, in several models (e.g. Tsotsos, 2011), binding depends upon filtering, and vice versa.

A key issue is whether these dependencies also hold when global control is involved. For example, although attentional binding may require filtering, does it require *attentional* filtering? Experimental results currently provide only limited guidance, in large part because of the practical difficulties involved. However, a case might be made for an *alignment thesis*: each attentional process requires attentional control of the contingently selective processes upon which it depends, so as to ensure that all the relevant processes operate on the appropriate inputs.

4. Relation to Visual Experience

Having defined a set of attentional processes, the next step is to examine how they might relate to conscious visual experience—or more precisely, to its representational content[9]. Consciousness is often believed to participate in the consolidation of information over a global scale (Baars, 1988; Cohen & Dennett, 2011; Dehaene & Changeux, 2011). As such, it could be involved in a given process in several ways. It could, for example, participate in the global *transmission* of information associated with that process, such as the information input or output—e.g., results that are broadcast to other areas. Or it could participate in the control of the *process itself*—e.g., override particular bindings made reflexively at early levels (cf., Libet, 1985). The extent to which conscious visual experience is involved in such functions appears to be a contingent matter, as is the extent to which these functions involve visual attention. Such issues have proven difficult to resolve (see e.g., De Brigard & Prinz, 2010; Kentridge, 2011).

A potential way to simplify this problem is to reduce it to parts that are each concerned with a particular aspect of conscious experience. Two types of consciousness are often distinguished: *P-consciousness* (phenomenal consciousness), involving the phenomenal aspects of experience, and *A-consciousness* (access consciousness), involving the representational aspects that can be reported verbally (Block, 1995). But given that different kinds of structure are created by different attentional mechanisms, another—possibly complimentary—set of distinctions might be drawn, based on the kinds of structures involved. Just as the experience of color and motion are distinct kinds (or at least aspects) of experience concerned with distinct *physical* properties of the world, so might there be kinds of experience concerned with distinct *structural* properties. Such distinctions might in turn be associated with differences in the properties of the experience itself (e.g., different time constants, due different underlying processes). If so, there may well result a clustering of properties that could allow various kinds of experience to be defined, and related to each other.

[9] Or more precisely yet—to the representational content of experience that is *sensory*, i.e., that results when our eyes are open and is used for getting around in the world. The phenomenal experience encountered in dreams or mental imagery, for example, is not considered here.

4.0. Attentional Sampling / Visual Uniformity

Information pickup—and thus, sampling—is necessary for any kind of visual perception. But although our visual experience can be vivid and compelling, it does not usually correspond directly to what is picked up at the retina. For example, although the eyes generally saccade several times a second (often over several degrees of visual angle), our subjective impression is of a single, stable "picture". In addition, variations across the eye in the range of colors and motions sampled (Section 3.0) are never part of this picture—a uniform resolution and range is experienced throughout. The mechanisms that create such stability and uniformity are largely unknown (Bridgeman, van der Heijden, & Velichkovsky, 1994; O'Regan & Noë, 2001), although they may depend in part upon the positional invariance established during attentional filtering (e.g., Tsotsos, 2011). If so, their involvement suggests that although attentional sampling (via the alignment thesis) may be necessary for visual experience, it cannot be sufficient.

It has been argued that the need for attentional sampling does not apply at a local level. For example, although the retina contains a *blind spot* where no photoreceptors exist (and thus no sampling occurs), what is experienced at that location appears to be “filled in” on the basis of adjacent information, so that the gap is not noticed (Ramachandran, 1992). However, the nature of such filling-in is problematic; it has been argued that there may actually be no experience of the missing information (see e.g., Dennett, 1992; Pessoa, Thompson, & Noë, 1998).

4.1. Attentional Filtering / Fragmentary Experience

Filtering—or at least, gating—appears necessary for conscious visual experience of any kind. In binocular rivalry, for example, observers fail to experience anything from the unmonitored eye, in which the output has likely been suppressed (see Rensink, 2009). A similar failure occurs in inattentional blindness when an unattended item smaller than a few degrees in size disappears completely (Mack & Rock, 1998), and in the neurophysiological condition of *neglect*, in which items in one hemifield are not seen at all (Bisiach, 1993). In all these conditions, gating has presumably shut down entirely.

Although it has been claimed that visual experience can occur without "attention" (Braun & Sagi, 1990; Koch & Tsuchiya, 2007), it may be that attentional gating is still required. In the dual-task design typically used in such experiments (Figure 3), observers given the task of identifying a pattern at the center of a display cannot also report the shape of an item located outside this zone. But they *can* report—and presumably experience—its color and orientation (Braun & Sagi, 1990; Fei-Fei, VanRullen, Koch, & Perona, 2002). Likewise, when observers track one event and do not recognize another in the background (suggesting a limit in holding or individuating), they occasionally report seeing "something" of the background, even though they cannot say what it is (Neisser & Becklen, 1975). Such reports suggest that what is limited in these dual-task conditions is not attentional processing in its entirety, but only attentional *binding*.

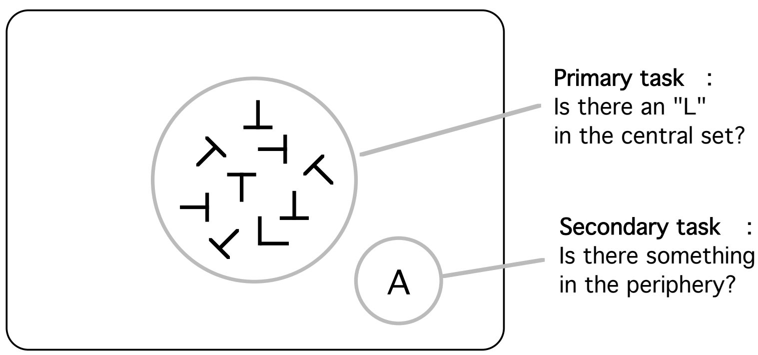


Figure 3. Example of dual-task design often used in investigations into how perception relates to attentional processing (e.g., Braun & Sagi, 1990; Fei-Fei et al., 2002).

A similar experience of localized properties without sophisticated structure is encountered during the perception of briefly-presented images. Observers can reliably determine the meaning (or *gist*) of an image presented for 30-60 milliseconds (Loschky & Larson, 2009; Thorpe, Fize, & Marlot, 1996); they can even detect the presence of animals under such conditions (Fei-Fei et al., 2002), as well as relatively abstract quantities such as the average size of a set of items (Chong & Treisman, 2003). The experience reported for such brief exposures in all these situations is one of a fleeting array of simple colors and shapes, with relatively little sophisticated structure (e.g., Fei-Fei, Iyer, Koch, & Perona, 2007).

Such results therefore suggest the possibility of a distinct kind of visual experience—*fragmentary experience*—in which the observer has access primarily to a dense array of simple localized features with little intrinsic structure[10] (Rensink, 2010; 2013); in some ways, it is similar to what is experienced when viewing an Impressionist painting. The content of such experience contains only those properties (such as color, motion, and orientation) that can be measured on the basis of local information, assigned to each point in the image, represented as scalars or low-dimensional vectors, and for which well-defined distance measures exist. The simplicity of these properties is consistent with the fact that they can be experienced in stimuli presented for as little as 30 milliseconds. Their simplicity is also consistent with the fact that in terms of phenomenal character, they are unproblematically visible, constituting the "raw stuff" of visual experience.

Fragmentary experience may be similar to the "background consciousness" thought to occur in iconic-memory displays, where contents are fleeting and appear to contain more than can be reported (Iwasaki, 1993). It may also include the "partial awareness" of word fragments experienced in brief displays (Kouider, de Gardelle, Sackur, & Dupoux, 2010), provided these are limited to simple components such as oriented line segments or localized structures (e.g., corners) formed by reflexive binding. Its focus on "raw" sensory qualities suggests that fragmentary experience may also be related to P-consciousness (Block, 1995), although it does not include non-local structures and shapes: if a set of blue patches, say, is arranged into a triangle, that triangle will not be explicitly represented—and thus not experienced—at this level. Note that fragmentary experience can be accompanied by abstract categorization (e.g., identification of scenes and animals) even if there is little or no conscious experience of the non-local structure of the underlying stimuli.

---

[10] This is referred to as *ambient experience* in Rensink (2013). An interesting issue is whether this is best viewed as a distinct kind of *experience* or an experience of a distinct kind of *structure*. In a similar vein: does someone suffering from integrative agnosia (who can experience only simple features) have an experience different in kind from that of most observers, or the same experience that is more restricted in its structure? Owing to the clustering of traits sketched here, it appears more appropriate to describe fragmentary experience as differing in kind. Similar considerations apply to the other kinds of experience described in the following sections.

Fragmentary experience often involves gating that extends over a relatively wide expanse, possibly allocated in a sequential way (VanRullen et al., 2004). As the primary task becomes more demanding, fragmentary experience of the scene—along with the perception of its gist—begins to fail (Cohen, Alvarez, & Nakayama, 2011). If this is because the extent of gating has been reduced to remove noise in order to facilitate the primary task (cf. Lavie, 1995), it would suggest that filtering—or more precisely, the information that filtering transmits—is needed for *any* process involving the fragmentary properties of an image, including the generation of fragmentary experience.[11] This is compatible with the suggestion that such filtering must be attentional in order to yield at least a fragmentary experience of a stimulus (Mack & Rock, 1998).

4.2. Attentional Binding / Assembled Experience

When given the task of identifying a pattern at the center of a briefly-presented display (Figure 3), observers can still detect simple fragmentary properties outside this zone, such as localized colors or orientations (Braun & Sagi, 1990; Braun, 1994). This has sometimes been considered evidence for visual experience without attention (Braun & Sagi, 1990). But given the distinctions made here, a more nuanced possibility arises: without attentional binding, fragmentary experience is still possible, but nothing more.

The kind of experience that *is* encountered when binding succeeds is that of simple sensory properties such as color and motion (fragmentary experience) along with a "layer" of static structure. Given its involvement with structure, this might be termed *assembled experience*. In some ways, it is similar to what is experienced under stroboscopic conditions, where brief flashes of light remove the information needed to perceive motion, yet still allow other basic properties—including form—to be seen. It can be encountered in displays presented for 100-150 milliseconds (Fei-Fei et al., 2007), the time needed for binding. Although structures bound preattentively (and thus, reflexively)can pre-empt immediate access to their components, their

[11] Given that filtering operates on representations that contain a degree of organization, it follows that what is experienced in a fragmentary way is not the raw information that enters the eye, but a quantity that has already been abstracted to some extent.

components can be accessed—and thus, experienced—by further (attentional) control of the binding process (Rensink & Enns, 1995).

In this view, any aspect of static organization is manifest only in assembled experience, which requires binding[12]. Note that for many aspects of binding (Section 3.2), local information does not suffice. In the case of shape, for example, although some properties can be defined locally (e.g., orientation or curvature at a point), others cannot (e.g., symmetry or size). Moreover, the explicit representation of bound structure involves associations, which cannot always be well represented by scalars or low-dimensional vectors[13]. Nor is there always a well-defined distance measure between structures. The structures experienced—such as shapes that extend over space—also differ from fragmentary properties in that they need not be dense (i.e., not present at each point in the image), but can be distinct elements that cover the visual field in a much sparser way. Finally, more sophisticated processing is likely involved; indeed, if the perception of shape is taken to be the beginning of concept formation (Arnheim, 1969), assembled experience may be the first stage (or level) where the perception of concepts occurs (cf. Prinz, 2006). It is likely that binding must be attentional if a viewer is to have an assembled experience of the stimulus (Treisman & Gelade, 1980).

The view of assembled experience as a combination of unstructured sensory properties and superimposed form is vaguely reminiscent of the hylomorphism of Aristotle, where substance is considered to be a compound of unstructured matter and superimposed form. But although assembled experience may contain form, it is not limited to this; other types of bound structure can also be experienced (e.g. a particular juxtaposition of two colors). Moreover, structure can be largely experienced on its own—e.g., the perception of a group that transcends the visible fragments it links. Assembled experience may be related to P-consciousness in that it includes the

---

[12] There may be sub-types of assembled experience, corresponding to sub-types of attentional binding (Section 3.2). In the interests of simplicity, this possibility is not discussed here.

[13] A shape could be represented as a scalar or vector under some conditions—e.g., a closed curve could be described in terms of its compactness. But unless there are tight constraints on the set of possible shapes, such a measure will capture only one aspect of its structure.

experience of both fragmentary properties and nonlocal shapes; the difference between the experience of sparse assembled structure and dense fragmentary properties may be related to the proposal that perceptual consciousness overflows cognitive access (Block, 2011). Like fragmentary experience, assembled experience can be accompanied by semantic categorization—e.g., identification of animal species (Fei-Fei et al., 2007). In such cases, the combination of raw sensory experience, structure, and semantic attribution might be considered a more "complete" form of static perception (cf. Prinz, 2006).

Meanwhile, given that attentional binding is needed for assembled but not fragmentary experience, the possibility arises of not just one type of inattentional blindness, but *two* (Figure 4):

- *Type 1*: a failure of *fragmentary* experience, caused by lack of *gating*
- *Type 2*: a failure of *assembled* experience, caused by lack of *binding*

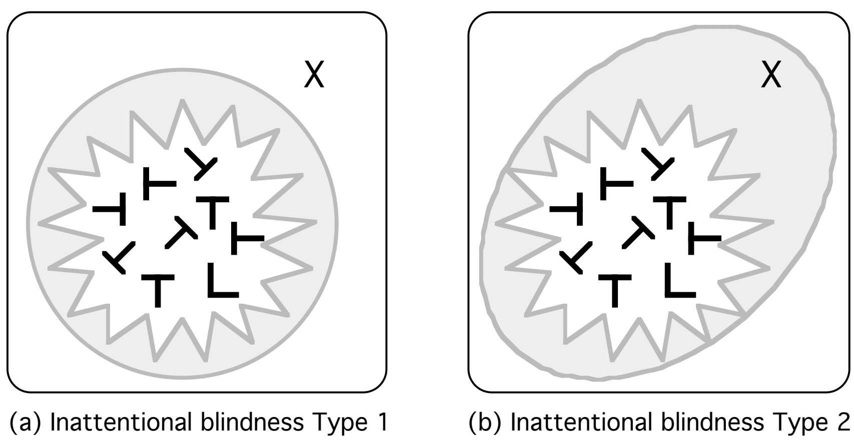


Figure 4. Types of inattentional blindness. (a) Type 1. Allocation of both attentional binding (jagged border) and filtering (smooth border) is limited to the central group, resulting in a failure to transmit any information about the "X". This causes a failure to experience the "X" even in a fragmentary way. (b) Type 2. Attentional filtering extends to the "X", but binding is still absent. This causes a failure to experience the "X" in an assembled way, although a fragmentary experience of it is still possible.

Empirical work suggests that Type 1 occurs only for items less than about 1°, at least when located in the fovea (see Rensink, 2013); it may be that gating can be shut down completely only if an item is in a single "gating zone". If so, Type 2 would be the only type of inattentional blindness encountered for larger stimuli (e.g., Neisser & Becklen, 1975; Simons & Chabris, 1999). This may explain the common belief that "inattentional blindness seems at odds with introspection" (Wolfe, 1999): since many tests involve relatively large stimuli, the observer might have inattentional blindness Type 2 while still having a fragmentary experience of the stimulus, since conditions would not be suitable for inattentional blindness Type 1.

4.3. Attentional Holding / Coherent Experience

In the same way that assembled experience involves static structure, another kind of experience could involve dynamic structure. This would include not only what is contained in fragmentary (and perhaps assembled) experience, but also an impression of *continuity*—of an object or event persisting over time. Such *coherent experience* is similar in some ways to the *object consciousness* proposed on the basis of verbal reports of structure, which likely involve visual short-term memory (Iwasaki, 1993). It may also connect to the idea of A-consciousness (Block, 1995), provided the latter concept applies to reporting the structure of objects or events that extend over time. Indeed, the apparent difference in capacities between this and fragmentary experience might account for the impression that an observer can report much less than what is experienced in a momentary glance (Block, 2011; see Kouider et al., 2010 for a somewhat similar proposal).

Attentional holding appears necessary for the coherent experience of an object (Rensink, 2002).[14] Whether it is also *sufficient* depends on what is meant by “object”. If this refers to the physical object, holding will generally be insufficient: relatively little information can be maintained in coherent form, preventing most properties of a physical object from being represented at any given time. But if "object" refers to the *representation* of the object (i.e., the corresponding visual object) that is always available for conscious report, then holding would be sufficient (Rensink, 2002; see also De Brigard & Prinz, 2010). Either way, given the necessity for

[14] In the view proposed here, change blindness—the failure to see change caused by a lack of attentional holding—could be described as "inattentional blindness Type 3".

attentional holding (and the complexity of the associated processes) is relatively time-consuming, this kind of experience taking up to 250-300 milliseconds to emerge.

As mentioned in Section 3.3, two kinds of continuity can be distinguished, based on whether the representation of an object is maintained throughout time, or whether it is recreated and matched with a trace in long-term memory, with differences then noted. These likely form the basis for different kinds of experience: *dynamic change*, where an object is experienced as dynamically transforming, and *completed change*, where the object is simply perceived as having changed at some point (Rensink, 2002). The former is encountered in coherent experience; the latter might involve a higher-level feeling of recognition.

Another kind of experience possibly related to these may underlie reports from some observers that they sometimes "sense" or "feel" a change in an image without having any clear idea of exactly what or where it is (Rensink, 2004). This *sensing* may be a distinct form of awareness involving some of the mechanisms that underlie coherent experience, although it likely also draws upon mechanisms that are different (Busch, Fründ, & Hermann, 2010).

4.4. Attentional Individuating / Perceptual Continuity

At a subjective level, individuating requires effort, consistent with it involving a form of attention (Scholl, 2009). It can also be accompanied by—at least during tracking—an experience of perceptual continuity somewhat similar to that experienced during holding. However, there is no experience here of the structure of the items individuated (e.g., their shape); for example, when an item is tracked, the only relevant property appears to be the position of its center of mass (Scholl, Pylyshyn, & Feldman, 2001).

It is unclear whether attentional individuating is associated with a distinct kind of experience—e.g., a continuity different in some way from that encountered in holding. Because individuating enables more effective control, its effects are likely to be exhibited mostly—if not entirely—via the facilitation of other kinds of attentional process. And if no new kinds of basic control are involved, it may be that no new kinds of conscious visual experience are needed. In any event, further progress on this issue will likely require additional empirical work to separate out the effects of individuating from those of the processes that it facilitates.

4.5. Nonattentional Processing / Dark Structure

Many visual processes appear to operate in the complete absence of visual experience (see e.g., Dehaene & Changeux, 2011). For example, in inattentional blindness Type 1, an unseen item can affect the treatment of a subsequently-presented item that is semantically related to it (Mack & Rock, 1998). Likewise, although observers generally fail to visually experience a target presented for an extremely brief duration (15-20 milliseconds), such an unseen item can speed up the identification of a subsequent item related to it (e.g., Naccache, Blandin, & Dehaene, 2002).

An important issue concerns the status of such unseen stimuli. One possibility is that their contents are not consciously experienced until some kind of attentional process operates on them; in analogy with "preattentive", these could be said to be *preconscious* (see, e.g., Dehaene & Changeux, 2011). However, there is also the possibility of *dark structure*: representational content (formed reflexively) that can *never* be part of conscious visual experience (Rensink, 2013). Dark structure might exist in various parts of the nervous system—e.g., the dorsal stream, which is believed to be exclusively concerned with action (see Goodale & Milner, 1992).

An interesting possibility is that conscious experience might participate in the control of dark structure to some extent—e.g., adjusting filter parameters or undoing binding that was created reflexively—with only the controlled aspects of the end results being consciously experienced. In the extreme version of this, conscious visual experience would essentially be a "control panel" that enables important aspects of fine-grained (and likely sophisticated) control, but does not participate in the bulk of visual processing, which would operate in the dark, as it were (cf. Norretranders, 1999). From this point of view, the conscious experience of a property would be inextricably linked with the control of its associated attentional processes. (If control of those processes were not sufficient for that kind of experience, it would imply that other, nonconscious forms of attentional control also exist.) Various forms of nonattentional processing, meanwhile, could still take place in the background.

4.6. Visual attention / Visual experience

Bringing together the points above, a pattern begins to emerge in the way that visual attention relates to visual experience (Figure 5). First, attentional filtering appears to be necessary for all three kinds of experience, and is the only attentional process needed for fragmentary experience. Next, attentional filtering and binding are necessary for assembled experience, while filtering, holding, and possibly binding are needed for coherent experience. (Because sampling is necessary for visual perception in general, its status is not discussed in detail here.) As such, the results of experimental work to date seem consistent with a simple principle: *for any kind of conscious experience, a distinct kind of attentional process is necessary*. (See also Cohen, et al., 2012). Consistent with the alignment thesis, the various kinds of attentional processes appear to form a cascade, with the more complex ones drawing upon the less complex. If so, this suggests that the different kinds of visual experience might correspond to *grades* that involve increasingly complex levels of structure.

	Filtering	Binding	Holding	Individuating
Coherent	Yes	?	Yes	No
Assembled	Yes	Yes	No	No
Fragmentary	Yes	No	No	No

Figure 5. Dependence of kinds of visual experience on kinds of attentional process. Each rectangle indicates whether the given kind of attentional process necessarily accompanies the given kind of experience. Status is based on data drawn from the sources discussed in Section 3 and Rensink (2013).

The extent to which each kind of attentional process suffices for visual experience is less clear. Holding may be sufficient for coherent experience, given that its outputs are visual objects

(Section 4.3). And since visual short-term memory is needed to form verbal reports, which are often used as an indicator of conscious processing (see e.g., Dehaene & Changeux, 2011), this could account for the proposal that attention—or more precisely, attentional holding—is both necessary and sufficient for conscious experience (De Brigard & Prinz, 2010). Meanwhile, filtering can be controlled in the absence of consciously-experienced stimuli (Kentridge, Nijboer, & Heywood, 2008), suggesting that while attentional filtering may be necessary for conscious (fragmentary) experience, it is not sufficient. There is also some evidence that binding can occur for unseen stimuli (see Kentridge, 2011), although it is unclear whether such binding is attentional or reflexive. More work is needed on these issues.

4.7. Co-ordination

Given the various kinds of visual experience posited here, how might these enter into our everyday experience of the world? It may be best to begin by considering how scene perception is believed to work. One proposal involves three systems co-ordinated so that attention—or more precisely, attentional holding—creates the right visual object at the right time; if this process is managed sufficiently well, the result is a *virtual representation* of the scene, which can be treated as if it were coherent everywhere (Rensink, 2000a). Such co-ordination can be also be applied *within* an object: even though only a small amount of information can be held in coherent form at any time, if this is done for the right part at the right time, the result would effectively be a coherent, detailed representation of the entire object (Rensink, 2001). An important consequence of this proposal is the constraint that while any visual representation can be dense (and volatile) or nonvolatile (and sparse), no visual representation can be both dense *and* nonvolatile.[15]

[15] The triadic architecture posited that the representation of dynamic structure in a scene was sparse. It also posited that the representation of static structure—experienced or not—could not be both dense and nonvolatile. But although the existence of change blindness can provide empirical evidence for assertions about dynamic structure, it was acknowledged that it cannot do so for static structure (e.g., Rensink, 2000b, p. 1475; Rensink, 2002, p. 266). This amounted to a claim that the attentional processes for dynamic structure differed from those for static. As such, it could be seen as the first step in the development of the framework presented here.

If several kinds of attentional processes exist, a similar kind of co-ordination might be possible among these, and thus, among the corresponding kinds (or grades) of visual experience. Assuming that each kind of experience aligns reasonably closely with its corresponding attentional processes[16], their distribution across space would be subject to particular constraints (Figure 6). To begin with, a zone of fragmentary experience would correspond to the area of spatial filtering. Given that filtering is necessary for binding and holding, zones of assembled and coherent experience would necessarily be located within the zone of fragmentary experience. Likewise, if attentional binding is needed for holding, the zone of coherent experience would need to be within the zone of assembled experience. Ideally, the intersection of all three zones would align with the center of fixation, so that the information contained within would have maximal resolution. Such an arrangement would allow coherent experience to remain relatively sparse in content and limited to the items held, while concurrently supporting other kinds of (static) experience with greater informational density or extent.[17]

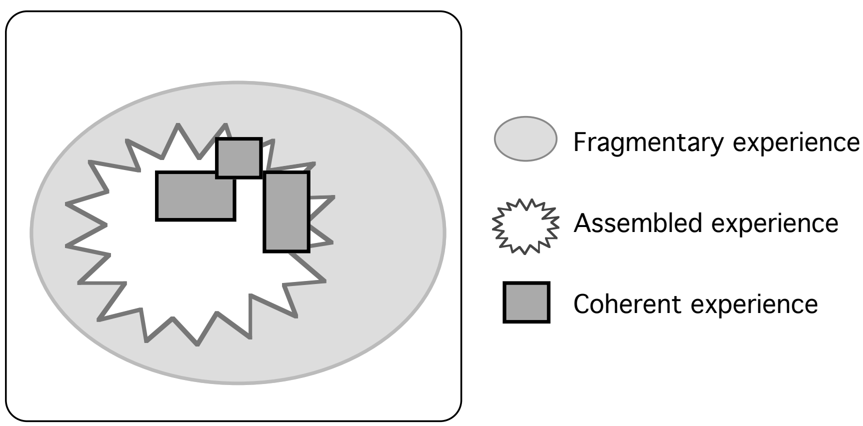


Figure 6. Zones of different kinds (grades) of visual experience. This assumes that each kind aligns reasonably closely with the extent of its corresponding attentional processes. The zone of fragmentary experience extends over the area gated by attentional filtering; assembled and coherent experience are located within it. If coherent experience involves bound structure, the zone of coherent experience would likewise exist within the zone of assembled experience.

[16] Even if an attentional process is necessary for its corresponding kind of experience, it may not be sufficient. As such, there may not be a 1:1 mapping between the distributions of the two.
[17] The extent of fragmentary and assembled experience is an open issue. But given that a large image can be experienced in a fragmentary way when presented briefly, fragmentary experience likely extends over a considerable area, at least when attentional filtering is diffuse.

If the appropriate attentional process could be applied to the appropriate part of the input at the appropriate time, it would create a *virtual experience* of all aspects of structure existing everywhere. It is only when such co-ordination breaks down—such as under the demanding conditions encountered in a controlled experiment on attentional allocation, or when a brain lesion interferes with control processes—that the separate components of visual experience would become evident.

5. Prospects

The relationship between visual experience and visual attention is a complex one. Determining its nature has proven to be a challenge, in terms of both the conceptual issues to be addressed and the empirical issues faced. This paper has explored one possible way of handling this challenge: fractionating both attention and visual experience into components, each concerned with a different kind of structure in the world. In this approach, the original issue of how attention relates to visual experience is replaced by a set of simpler issues, each concerned with how a particular kind of attentional process relates to a particular kind of visual experience. As the discussion here has shown, such an approach is not only feasible and helps cast light on the relationship between attention and visual experience, but also raises new issues, such as the existence of different kinds of inattentional blindness. As such, it appears to be worth developing further.

Of course, such an attempt is at heart a gamble: nature may simply not be this way. But even if this approach does not succeed entirely, parts may still be helpful. For example, regardless of any connection with visual experience, it may still be useful to characterize "paying attention" as the co-ordinated control of several processes, each concerned with a particular aspect of structure. And if the particular groupings proposed here do not turn out to capture reality sufficiently well, the principles suggested here as the basis of a taxonomy might still help create an improved classification.

Similar prospects exist for the proposal that conscious visual experience might be fractionated into distinct components (or possibly, grades). If nothing else, this view raises several interesting questions: To what extent is the experience of color of the same "kind" of experience as that of motion? Is the experience of continuity over time "perceptual" in the same way as the experience of the color blue? Similar considerations apply regarding the extent to which conscious visual experience is a virtual phenomenon: What kinds of co-ordination exist? How might these break down? And even if visual experience does turn out to match our impressions and actually be a unitary phenomenon, there would still be great value in knowing this, and knowing why this should be so.

Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council of Canada. Many thanks to Paul Coates, Carolyn Dicey Jennings, Minjung Kim, and John Tsotsos for their helpful comments on earlier versions of this paper. Thanks also to Paul Coates for providing this vision scientist with an opportunity to interact with an interesting group of philosophers; I hope that in return they find something of interest here.

References

Allport, A. (1989). "Visual attention." In M.I. Posner (Ed.), *Foundations of Cognitive Science* (pp. 631-682). (Cambridge, MA: MIT Press).

Allport, A. (1993). "Attention and control: Have we been asking the wrong questions? A critical review of twenty-five years." In D.E. Meyer & S. Kornblum (Eds.), *Attention And Performance XIV: Synergies In Experimental Psychology, Artificial Intelligence, And Cognitive Neuroscience* (pp. 183-218). (Cambridge, MA: MIT Press).

Arnheim, R. (1969). *Visual Thinking* (ch. 2). (Berkeley CA: University of California Press).

Baars, B.J. (1988). *A Cognitive Theory of Consciousness*. (Cambridge: University Press).

Bahrami, B. (2003) "Object property encoding and change blindness in multiple object tracking." *Visual Cognition*, *10*: 949–963.

Ballard, D.H., Hayhoe, M.M., Pook, P.K., & Rao, R.P.N. (1997). "Deictic codes for the embodiment of cognition." *Behavioral and Brain Sciences*, *20*: 723-767.

Barlow, H.B., & Mollon, J.D. (Eds.). (1982). *The Senses*. (Cambridge: University Press).

Bartolomeo, P., & Chokron, S. (2002). "Orienting of attention in left unilateral neglect." *Neuroscience and Biobehavioral Reviews, 26*: 217-234.

Bisiach, E. (1993). "Mental representation in unilateral neglect and related disorders: The twentieth Bartlett Memorial Lecture." *Quarterly Journal of Experimental Psychology*, *46*: 435-561.

Block, N. (1995). "On a confusion about a function of consciousness". *Behavioral and Brain Sciences*, *18*: 227-247.

Block, N. (2011). "Perceptual consciousness overflows cognitive access." *Trends in Cognitive Sciences*, *15*: 567-575.

Bridgeman, B., van der Heijden, A.H.C., & Velichkovsky, B.M. (1994). "A theory of visual stability across saccadic eye movements." *Behavioral and Brain Sciences*, *17*: 247-292.

Braun, J. (1994). "Visual search among items of different salience: Removal of visual attention mimics a lesion in extrastriate area V4." *Journal of Neuroscience*, *14*: 554-567.

Braun, J., & Sagi, D. (1990). "Vision outside the focus of attention." *Perception & Psychophysics*, *48*: 45-58.

Broadbent, D.E. (1982). "Task combination and selective intake of information." *Acta Psychologica, 50*: 253-290.

Busch, N.A., Fründ, I, & Hermann, C.S, (2010). "Electrophysiological evidence for different types of change detection and change blindness."*Journal of Cognitive Neuroscience*. *22*: 1852-1869.

Carpenter, R.H.S. (1988). *Movements of the Eyes* (2nd ed.). (London: Pion).

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

Chun, M.M. (1997). "Types and tokens in visual processing: A double dissociation between the attentional blink and repetition blindness." *Journal of Experimental Psychology: Human Perception and Performance, 23*: 738-755.

Chun, M.M., Golomb, J.D., & Turk-Browne, N.B. (2011). "A Taxonomy of External and Internal Attention." *Annual Review of Psychology*, *62*: 73-101.

Cohen, M.A., Alvarez, G.A, & Nakayama, K. (2011). "Natural-Scene Perception Requires Attention." *Psychological Science, 22*: 1165-1172.

Cohen, M.A., Cavanagh, P., Chun, M.M., & Nakayama, K. (2012). "The attentional requirements of consciousness." *Trends in Cognitive Sciences*, *16*: 411-417.

Cohen, M.A., & Dennett, D.C. (2011). "Consciousness cannot be separated from function." *Trends in Cognitive Sciences*, *15*: 358-364.

Crowne, D.P. (1983). "The frontal eye field and attention." *Psychological Bulletin*, *93*: 232-260.

De Brigard, F., & Prinz, J. (2010). "Attention and consciousness." *Wiley Interdisciplinary Reviews: Cognitive Science, 1*: 51–59.

Dehaene, S., & Changeux, J.-P. (2011). "Experimental and theoretical approaches to conscious processing." *Neuron*, *70*: 200-227.

Dennett, D.C. (1992). "'Filling in' versus finding out: a ubiquitous confusion in cognitive science." In H.L. Pick Jr, P. van den Broek, & D.C. Knill (Eds.), *Cognition: Conceptual and Methodological Issues* (pp. 33-49). (Washington DC: American Psychological Association).

Driver, J., Davis, G., Russell, C., Turatto, M., & Freeman, E. (2001). "Segmentation, attention and phenomenal visual objects." *Cognition, 80*: 61-95.

Farah, M.J. (2004). *Visual Agnosia* (2nd ed.). (Cambridge MA: MIT Press).

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). "What do we perceive in a glance of a real-world scene?" *Journal of Vision*, *7, 10:* 1-29.

Fei-Fei, L., VanRullen, R., Koch, C., & Perona, P. (2002). "Rapid natural scene categorization in the near absence of attention." *Proceedings of the National Academy of Sciences USA*, *99*: 9596-9601.

Franconeri, S.L. (2013). "The nature and status of visual resources." In D. Reisberg (Ed.), *Oxford Handbook of Cognitive Psychology* (pp. 147-162). (New York: Oxford University Press).

Goodale, M.A., & Milner, A.D. (1992). "Separate visual pathways for perception and action." *Trends in Neurosciences, 15: 20*-22.

Hatfield, G. (1998). "Attention in early scientific psychology." In R.D. Wright (Ed.), *Visual Attention* (pp. 3-25). (Oxford: University Press).

Henderson, J.M., & Hollingworth, A. (1998). "Eye movements during scene viewing: An overview." In G. Underwood (Ed.), *Eye Guidance in Reading and Scene Perception* (pp. 269-293). (Oxford: Elsevier).

Itti, L., Rees, G., Tsotsos, J.K. (2005). *The Neurobiology of Attention*. (Burlington MA: Academic).

Iwasaki, S. (1993). "Spatial attention and two modes of consciousness." *Cognition*, *49*: 211-233.

Jennings, C.D. (2012). "The subject of attention." *Synthese*, *189*: 535-554.

Jensen, M.S., Yao, R., Street, W.N., Simons, D.J. (2011). "Change blindness and inattentional blindness." *Wiley Interdisciplinary Reviews: Cognitive Science*, 2: 529–546.

Juan, C.-H., Shorter-Jacobi, S.M., & Schall, J. (2004). "Dissociation of spatial attention and saccade preparation." *Proceedings of the National Academy of Sciences USA*, *101: 15541-15544.*

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

Kanwisher, N., Yin, C., & Wojciulik, E. (1999). "Repetition blindness for pictures: Evidence for the rapid computation of abstract visual descriptions." In V. Coltheart (Ed.), *Fleeting Memories*: *Cognition Of Brief Visual Stimuli* (pp. 119-150). (Cambridge MA: MIT Press).

Kentridge, R.W. (2011). "Attention without awareness: A brief review." In Mole, C., Smithies, D., & Wu, W. (Eds.), *Attention: Philosophical and Psychological Essays* (pp. 228-246). (Oxford: University Press).

Kentridge, R.W., Nijboer, T.C.W., & Heywood, C.A. (2008). "Attended but unseen: Visual attention is not sufficient for visual awareness." *Neuropsychologia*, *46*: 864-869.

Koch, C., & Tsuchiya, N. (2007). "Attention and consciousness: two distinct brain processes." *Trends in Cognitive Sciences*, *11*: 16-22.

Kouider, S., de Gardelle, V., Sackur, J., & Dupoux, E. (2010). "How rich is consciousness? The partial awareness hypothesis." *Trends in Cognitive Sciences*, *14:* 301-307.

Krauzlis, R.J. (2005). "The control of voluntary eye movements: New perspectives." *The Neuroscientist, 11*: 124-137.

Lamme, V.A.F. (2003). "Why visual attention and awareness are different." *Trends in Cognitive Sciences***,** *7:* 12-18.

Lavie, N. (1995). "Perceptual load as a necessary condition for selective attention." *Journal of Experimental Psychology: Human Perception and Performance*, *21*: 451-468.

Libet, B. (1985). "Unconscious cerebral initiative and the role of conscious will in voluntary action." *Behavioral and Brain Sciences*, *8*: 529–66.

Loschky, L.C., Larson, A.M. (2009). "The natural/man-made distinction is made before basic-level distinctions in scene gist processing." *Visual Cognition, 18*: 513-536.

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

Maddox, W.T., Ashby, F.G, & Waldron, E.M. (2002). "Multiple attention systems in perceptual categorization." *Memory & Cognition*, *30*: 325-329.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (pp. 8-38). (San Francisco: W.H. Freeman).

Naccache, L. Blandin, E., & Dehaene, S. (2002). "Unconscious Masked Priming Depends on Temporal Attention." *Psychological Science, 13*: 416-424.

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

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

Norretranders, T. (1999). *The User Illusion: Cutting Consciousness Down to Size*. (New York: Penguin Books).

O'Regan, J.K., & Noë, A. (2001.) "A sensorimotor account of vision and visual consciousness." *Behavioral and Brain Sciences*, *24*: 939-73.

Pashler, H.E. (1999). *The Psychology of Attention*. (Cambridge MA: MIT Press).

Pessoa, L., Thompson, L., & Noë, A. (1998). "Finding out about filling in: a guide to perceptual completion for visual science and the philosophy of perception." *Behavioral and Brain Sciences, 21*: 723-802.

Posner, M.I., Walker, J.A., Friedrich, F.J., & Rafal, R.D. (1984). "Effects of parietal injury on covert orienting of attention." *Journal of Neuroscience, 4:* 1863-1874.

Prinz, J.J. (2006). "Beyond appearances: The content of sensation and perception." In T.S. Gendler & J. Hawthorne (Eds.), *Perceptual Experience* (pp. 434-460). (Oxford: University Press).

Pylyshyn, Z.W. (2003). *Seeing and Visualizing: It's Not What You Think*. (pp. 201-279). (Cambridge MA: MIT Press).

Ramachandran, V.S. (1992). "Filling in gaps in perception: Part I." *Current Directions in Psychological Science*, *1:* 199-205.

Rayner, K., Smith, T.J., Malcolm, G.L., & Henderson, J.M. (2009). "Eye movements and visual encoding during scene perception." *Psychological Science*, *20*: 6-10.

Rensink RA. (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. (2001). "Change blindness: Implications for the nature of attention." In M.R. Jenkin and L.R. Harris (Eds.), *Vision and Attention* (pp. 169-188). (New York: Springer.)

Rensink, R.A. (2002). "Change detection." *Annual Review of Psychology*, *53:* 245-277.

Rensink, R.A. (2004). "Visual sensing without seeing." *Psychological Science*, *15:* 27-32.

Rensink, R.A. (2009). "Attention: Change blindness and inattentional blindness." In W. Banks (Ed.), *Encyclopedia of Consciousness,* Vol 1 (pp. 47-59). (New York: Elsevier).

Rensink, R.A. (2010). "Seeing seeing." *Psyche*, *16*: 68-78.

Rensink, R.A. (2013). "Perception and attention." In D. Reisberg (Ed.), *Oxford Handbook of Cognitive Psychology* (pp. 97-116). (New York: Oxford University Press).

Rensink, R.A., & Enns, J.T. (1995). "Preemption effects in visual search: Evidence for low-level grouping." *Psychological Review, 102:* 101-130.

Riddoch, M.J. Humphreys, G.W. (1987). "A case of integrative visual agnosia." *Brain*, *110*: 1431-1462.

Robertson, L.C., & Kim, M.-S. (1999). "Effects of perceived space on spatial attention." *Psychological Science*, *10:* 76-79.

Scholl, B.J. (2009). "What have we learned about attention from multiple-object tracking (and vice versa)?" In D. Dedrick & L. Trick (Eds.), *Computation, Cognition, and Pylyshyn* (pp. 49-77). (Cambridge MA: MIT Press).

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.

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

Sperling, G., & Dosher, B.A. (1986). "Strategy and optimization in human information processing." In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of Perception and Human Performance, Vol 1* (pp. 1-65). (New York: Wiley).

Thorpe, S., Fize, D., & Marlot, C. (1996). "Speed of processing in the human visual system." *Nature, 381*: 520-522.

Treisman, A.M. (1969). "Strategies and models of selective attention." *Psychological Review*, 76: 282-299.

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

Tsotsos, J.K. (2011). *A Computational Perspective on Visual Attention*. (Cambridge, MA: MIT Press).

Ullman, S. (1984). "Visual routines." *Cognition, 18:* 97-159.

VanRullen, R., Reddy, L., & Koch, C. (2004). "Visual search and dual tasks reveal two distinct attentional resources." *Journal of Cognitive Neuroscience*, *16:* 4-14.

Wolfe, J.M. (1999). "Inattentional amnesia." In Coltheart, V. (Ed.), *Fleeting Memories: Cognition of Brief Visual Stimuli* (pp. 71-94). (Cambridge, MA: MIT Press).

Wright, R.D. (Ed.). (1998). *Visual Attention*. (Oxford: University Press).