16.

ENSEMBLE PERCEPTION: SUMMARIZING THE SCENE AND BROADENING THE LIMITS OF VISUAL PROCESSING

Jason Haberman and David Whitney

A vast amount of information is available to the visual system at any given moment. Processing the billions of bits of information on the retina is a monumental challenge, but the way in which the brain accomplishes this is not via brute force. Numerous studies have unequivocally demonstrated that the brain has severe processing limitations, resulting in a sparse representation of our environment (e.g., visual short term memory, change blindness, multiple object tracking, attentional blink; Luck & Vogel, 1997; Raymond, Shapiro, & Arnell, 1992; Rensink, O'Regan, & Clark, 1997; Scholl & Pylyshyn, 1999; Simons & Chabris, 1999). To surmount some of these limitations and lighten the computational load, the brain utilizes numerous heuristics. Such heuristics or assumptions are learned over time because of the predictability and stability of the visual world, and they work well enough. Rather than generating high fidelity representations of everything within our field of view, our brains exploit the world's statistical regularity to condense information. The leaves of a tree, the blades of grass, the tiles of the floor are redundant, giving rise to the percept of "tree-ness," "lawn-ness," and "floor-ness," respectively. The individual components of those textures are lost in favor of a concise, summary statistical representation.

The concept of summary representation has recently generated significant interest and debate within the vision science community (Alvarez & Oliva, 2008, 2009; Alvarez, 2011; Ariely, 2001, 2008; accompanying paper; Chong & Treisman, 2005a, 2005b; de Fockert & Marchant, 2008; Haberman & Whitney, 2007, 2009; Koenderink, van Doorn, & Pont, 2004; Myczek & Simons, 2008; Simons & Myczek, 2008). Sometimes also called ensemble coding or ensemble perception, summary representation refers to the idea that the visual system naturally represents sets of similar items (such as blades of grass) using summary statistics. Such a system is intuitively appealing and has far-reaching implications. Chong and Treisman (2003, accompanying paper), and, more recently, we (Haberman & Whitney, 2009) and other authors even suggest that summary representation can provide coarse information from sources across our entire visual field, driving the compelling impression that we have a complete and accurate picture of our visual world (accompanying paper; Haberman & Whitney, 2009). Thus, the "Grand Illusion" (Noe, Pessoa, & Thompson, 2000) may not be an illusion at all, but rather a noisy summary representation of all that we survey. Many of the individual details of a scene are inaccessible, but the "gist" is ever present. Ensemble coding serves as a computationally inexpensive means of obtaining valuable information about a scene. Put another way, ensemble coding may provide a viable algorithm that drives gist perception.

The concept of ensemble representation is not a new one. Aristotle described perception as a "mean" of sensory inputs, which could be used to identify stimulus changes as the "sense organ" gathered more information. Extensive psychophysical work since the 1980s has demonstrated that the visual system averages position (Morgan & Glennerster, 1991; Morgan, Watamaniuk, & McKee, 2000), direction of motion (Watamaniuk, Sekuler, & Williams, 1989; Williams & Sekuler, 1984), speed (Watamaniuk & Duchon, 1992), and orientation (Dakin & Watt, 1997; Morgan, Chubb, & Solomon, 2008; Motoyoshi & Nishida, 2001; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). The neural mechanisms that subserve ensemble perception may be straightforward. Perceiving the average direction of motion from a set of moving dots (or blowing snow), for example, is consistent with established physiological mechanisms of motion perception (Britten & Heuer, 1999; Britten, Shadlen, Newsome, & Movshon, 1992; Newsome & Pare, 1988); information may be pooled across low-level motion detectors operating in parallel, potentially obviating the involvement of serial attention (Watamaniuk & McKee, 1998; but see also Bulakowski, Koldewyn, & Whitney, 2007). Rather than perceiving each moving dot individually, the dominating percept is the average direction of motion.

More recent—and controversial—work has suggested that humans also derive a summary representation for the *size* of a set of arbitrary objects (Ariely, 2001; Chong & Treisman, 2003, accompanying paper; Chong & Treisman, 2005a, 2005b), and that this summary representation is favored over a representation of the individual items composing the set. Some researchers argue that perceiving the average size is a parallel process (Ariely, 2001; accompanying paper), similar to distinguishing two textures. This raises several interesting questions, including: Are there low-level feature detectors designed to operate on object size in a manner akin to motion or orientation? If not, how does average-size perception, if it is indeed parallel, bypass traditional limitations of serial attention? Are there other examples of ensemble coding that extend beyond low-level stimuli (i.e., motion, orientation, size)?

Although open questions remain (some of which are addressed later), it is clear that ensemble coding is connected to several areas of vision science, and this, in part, explains the growing interest in summary statistical perception. In addition to providing a window on "gist," ensemble perception has implications for the way we understand visual search, texture, depth, scene perception, memory, object recognition, and spatial vision. Because of its far-reaching and potentially controversial implications, research on ensemble perception is rapidly expanding. The remainder of this chapter surveys the history of this subfield, discusses ongoing debates, highlights in greater detail some of the more influential work, and speculates about where future work should be directed.

SURVEY OF SUMMARY STATISTICAL PERCEPTION

Although it was not always referred to as ensemble or summary statistical perception, this phenomenon has implicitly been examined in some form since the early twentieth century. Gestalt grouping (Wertheimer, 1923) may be viewed as an early conceptualization of summary representation. The gestaltists viewed emergent object perception as a synergy of lower-level inputs; the final percept was more than the sum of its parts. Researchers argued that the grouped object was the favored percept, and that the individual features were (at worst) lost or (at best) difficult to perceive (Koffka, 1935). Although gestaltists outlined several basic heuristics by which the visual system groups features (similarity, proximity, common fate, etc.), the underlying mechanism(s) driving this grouping, as well as the algorithm that supports it, remained elusive. It may be that gestalt grouping amounts to a summary statistical representation, and the mechanism of ensemble coding may provide an explanation for several gestalt phenomena.

Although gestalt phenomenology helped to define some elemental principles of object perception, researchers in this area were not explicitly thinking in terms of ensemble perception or summary statistical representation. Some of the earliest explicit work on ensemble coding was done from a social psychology perspective. In an extensive line of research, Norman Anderson outlined a simple yet flexible model called "integration theory" (Anderson, 1971). His work demonstrated that a weighted mean more precisely captured how information is integrated than a summation model. For example, subjects rated another individual more favorably when that person was described by two extremely positive terms compared to when that person was described by two extremely positive terms in addition to two moderately positive terms (Anderson, 1965). Anderson cited this as evidence that humans employ a weighted average when evaluating a complex situation (the weighting of a given descriptor could be influenced by any number of factors). If subjects were summing information, then the ratings for an individual, described by four positive terms (two extreme and two moderate), should have been higher than for the individual described by two positive terms (two extreme). Instead, the moderately positive terms pulled the overall impression down. Integration theory was extended to numerous other social contexts, including "group attractiveness" (Anderson, Lindner, & Lopes, 1973), shopping preferences (Levin, 1974), and even the perceived "badness" of criminals accused of certain crimes (Leon, Oden, & Anderson, 1973). Thus, it appears humans readily integrate semantic as well as social information, although the mechanism behind this process remains largely unknown. The implication is clear, however: social perceptions and attitudes may hinge on the same sort of underlying computations and mechanisms that allow us to perceive average orientation and direction of motion.

There is a substantial body of psychophysical work demonstrating integration or ensemble coding of low-level feature information, the mechanisms of which are fairly well understood. For example, humans precisely perceive the average direction of motion of a group of dots moving along unique local vectors (Watamaniuk et al., 1989; Williams & Sekuler, 1984). This summary is extracted in parallel using receptive fields dedicated to processing motion across the retina (Britten & Heuer, 1999; Frechette et al., 2005; Jancke, 2000). Similar averaging principles hold true across other low-level domains as well, including speed (Watamaniuk & Duchon, 1992), orientation (Dakin & Watt, 1997; Parkes et al., 2001), number (Burr & Ross, 2008), position (Alvarez & Oliva, 2008; Morgan & Glennerster, 1991), size (Ariely, 2001; Chong & Treisman, 2003, accompanying paper; Chong & Treisman, 2005b), and even shadows (Koenderink et al., 2004; Sanders, Haberman, & Whitney, 2008)—a testament to the elemental and far-reaching implications of ensemble coding. Given its flexibility, it may come as no surprise that summary representation extends to complex, higher-level objects, including faces (figure 16.1; de Fockert & Wolfenstein, 2009; Fischer & Whitney, 2011; Haberman, Harp, & Whitney, 2009; Haberman & Whitney, 2007, 2009; Sweeny, Grabowecky, Paller, & Suzuki, 2009).

PERCEIVING AVERAGE SIZE

Although it is very important to recognize the significance of the earlier work on feature averaging, the bulk of this chapter will focus on ensemble coding within the last decade. This is partly due to the fact that the vast majority of "ensemble perception" research published prior to then was done from a low-level psychophysics perspective. Although this research remains crucially informative, the importance of ensemble coding extends to attention, cognition, and several other broader areas of psychology. The growing interest in the field within the last decade was sparked by both its general appeal and because of the implications it held for traditional notions of perception and awareness.

The current era in the study of ensemble statistics began when Ariely (2001) provided evidence that observers could derive the average size from a set of dots varying in size. In fact, this summary representation was the favored representation, as observers implicitly extracted the mean information. Observers viewed sets of dots for two seconds, and then responded about whether a subsequently viewed test dot was a member of the set. The striking aspect of these data was not

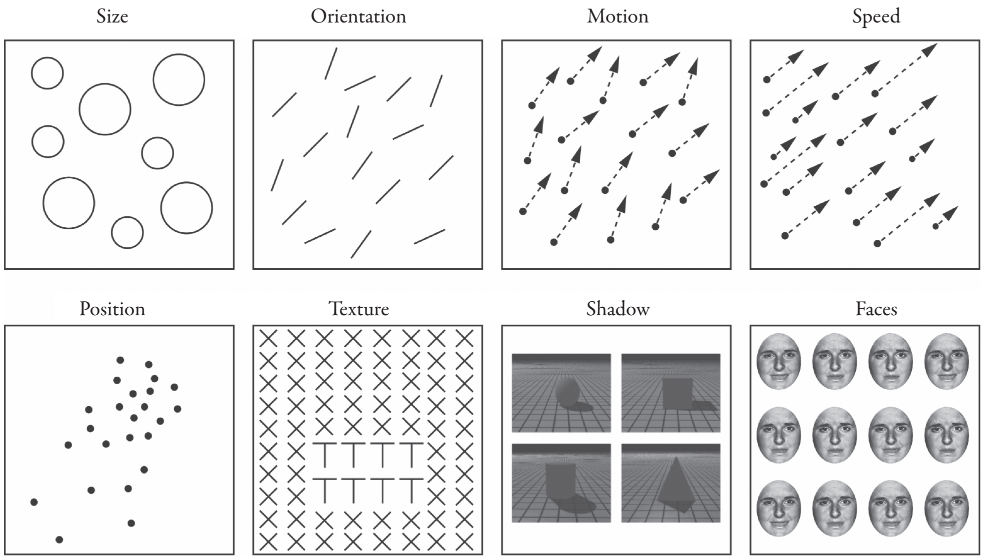

Figure 16.1 The various domains in which summary statistical representation occurs. The flexibility of summary representation suggests that it occurs across multiple levels along the visual hierarchy.

that observers performed poorly at the member identification task. As the size of the test dot approached the average size of the array of dots, observers were much more likely to respond that the test dot was a member of the set. Even though observers were instructed to attend to the individual members, they instead represented the summary of the set constituents. When explicitly asked about the mean size of a set of dots, observers were nearly as precise in discriminating the mean size of several dots as they were in discriminating the size of a single dot. Interestingly, mean discrimination performance seemed invariant to set size (up to 16 items), possibly suggesting that serial attentional mechanisms may not be required.

AVERAGE SIZE AND THE ROLE OF ATTENTION

Although ensemble representations had been established for many low-level visual attributes, the notion that these representations might be available effortlessly and underlie much of our subjective visual awareness of scene "gist" remained untested. Chong and Treisman (2003, accompanying paper) were among the first to systematically make the connection between ensemble representations and visual attention (Chong & Treisman, 2003, accompanying paper; Chong & Treisman, 2005a, 2005b). Based on several pieces of evidence, Chong and Treisman made a strong case that average size perception occurs "automatically" (although they qualified that term by describing ensemble coding as occurring "in parallel and without intention rather than without attention"; Chong & Treisman, 2005b). Robust average size perception was demonstrated across an impressive array of manipulations, showing immunity to changes in presentation (simultaneous versus successive), duration, and set distribution. In one series of experiments (accompanying paper), observers were asked to identify which of two sets of 12 circles had the larger mean size. The sets were either presented simultaneously (side-by-side) or successively. Observers' discrimination of the average size of the set was nearly as precise as their discrimination of the size of a single circle, regardless of presentation condition (simultaneously or successively). Equally important was the fact that set exposure duration had only a minor impact on average size discrimination. Even when the sets were presented for only 50 milliseconds, observers were able to derive an accurate estimate of the mean size.

Chong and Treisman (2003, accompanying paper) also attempted to address alternative strategies that observers might use in assessing the average size. In the initial experiments, the distribution of circle sizes used was uniform in a given set; that is, each size was equally represented. Observers could theoretically have compared the largest circle in each of the two sets to arrive at the correct answer, bypassing a mean calculation altogether. To control for this, Chong and Treisman tested multiple circle size distributions, including a "two-peaks" distribution, a normal distribution, a homogeneous distribution, and the original uniform distribution of circle sizes (figure 16.2a). Observers saw two sets of circles presented simultaneously. Both arrays either had the same distributions or different distributions. Performance on mean-size discrimination when viewing two identical distributions was as good as it was when viewing the original uniform distributions. Although performance was significantly worse when the simultaneously presented distributions differed, this amounted to a difference in discrimination ability of only 2 percent. According to the authors, this is evidence that participants were indeed averaging, since strategies of comparing single dots across sets would not have succeeded when viewing differing distributions. This seems reasonable, as the negligible difference in performance could be attributed to differences in variance (Callaghan, 1984, 1989; Dakin & Watt, 1997; Duncan & Humphreys, 1989; Morgan et al., 2008) introduced by manipulating the shape of the distribution of circles in each set.

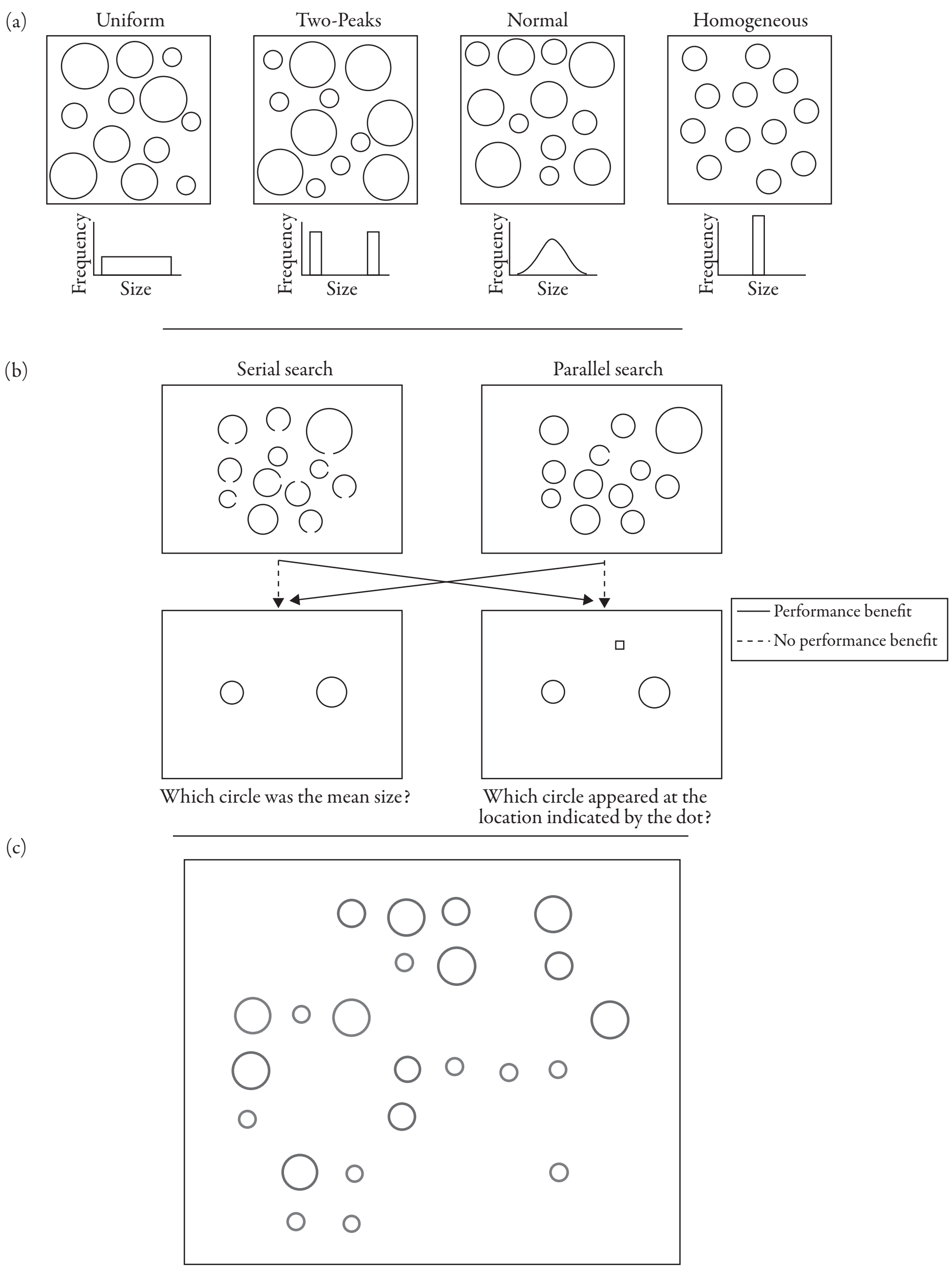

Figure 16.2 Some of the average size paradigms implemented by Chong and Treisman. **a.** By manipulating the distribution of circle sizes within the set, one can rule out potentially confounding strategies, such as examining only the largest circle size in each set. **b.** The dual task was used to examine the effects of attentional modulation on average size representation. In the serial search (searching for a closed circle among open ones), which required more focused attention, observers performed better in identifying whether a particular circle appeared at the indicated location. In the parallel search (searching for an open circle among closed ones), a task requiring more global attention, observers were better able to extract the average size of the whole set. Thus, one's attentional state can modulate summary statistical representation performance. **c.** Observers had to determine the average size of the color-defined set of circles given either a precue or a postcue. Remarkably, they were just as precise when they were given a postcue as when they were given a precue. Thus, observers could generate a mean representation for both sets simultaneously and without cost.

Although these thorough experiments suggest that ensemble coding occurs implicitly and perhaps in parallel, there is still evidence to suggest that performance is affected by manipulations of attention (Chong & Treisman, 2005a). Using an interesting technique, Chong and Treisman (2005a) asked observers to perform a dual task, which, on any given trial, included searching for a target (hard or easy search) and either a mean discrimination task or a member identification task (figure 16.2b). In the easy search task, observers had to find an open circle (a C) among a sea of closed circles, which the authors argued corresponded to a distributed or global mode of attention. In the hard search task, observers had to find the converse; a closed circle among an array of open circles (C's in various orientations), a task that putatively required focused or local attention (Treisman & Gormican, 1988). Following the search task, observers either judged which of two circles corresponded to the mean of the previously viewed set (mean discrimination) or which of two circles corresponded to a specific member of the previously viewed set (the location of which was indicated by a dot; membership identification). They hypothesized that engaging mechanisms of global attention should facilitate mean discrimination performance, and mechanisms of local attention should facilitate membership identification. Indeed, the results confirmed this (solid arrows in figure 16.2b), showing the predicted interaction between attentional mode and mean/member judgment.

Chong and Treisman (2005a) also found an attentional modulation in a second experiment that did not require observers to make judgments about the set constituents. Instead, observers made orientation judgments about a large rectangle that encompassed the array of circles (global attention) or a small rectangle in the center of the array of circles (local attention). The authors speculated that attending to the large rectangle would facilitate distributed attention, resulting in more precise average size representation than when attending to the small rectangle. Indeed, observers had superior mean discrimination performance when they had to judge the orientation of the large rectangle. Thus, even though ensemble coding can occur implicitly (Ariely, 2001; accompanying paper), the precision of mean representation (at least for size) does depend upon the spread of spatial attention.

Although summary statistical perception of size is modulated by attention, another elegant experiment (Chong & Treisman, 2005b) further demonstrated its automaticity. Observers discriminated the average size of a subset of an array of circles that was segregated from the rest of the array by color (figure 16.2c). Observers were remarkably precise on this task. In fact, average size perception did not differ whether the color cue preceded or followed the array of circles, and was no worse even when only a single color was presented. Because the cue did not make a difference in performance, the authors argue that the average size is computed automatically and across multiple sets, preceding or perhaps bypassing limitations imposed by the attentional bottleneck. They suggest that object binding is not necessary to extract a mean, and that, instead, a strategy akin to guided search may play an important role (Wolfe, Cave, & Franzel, 1989). If the visual system can rapidly (i.e., in parallel) segregate the scene using a feature map (in this case color), then average size representation should not be compromised by the presence of an irrelevant subset of circles. Although the role of attention in average size representation is an ongoing debate (Ariely, 2008; Chong, Joo, Emmanouil, & Treisman, 2008; Myczek & Simons, 2008; Simons & Myczek, 2008), as discussed later, these studies provide support for the existence of an automatic mechanism responsible for average size computation.

PERCEIVING AVERAGE ORIENTATION

The role of attention in average orientation perception is less controversial than it is in average size perception. There is both psychophysical and physiological evidence suggesting that average orientation representation is a parallel process. Some of the strongest evidence for this comes from Parkes and colleagues (2001), who showed that the orientation of a Gabor patch crowded out of awareness (i.e., observers were unable to discriminate its orientation) nonetheless influenced the perceived average orientation of an entire set of surrounding Gabor patches. Even though observers could not consciously individuate or scrutinize the target Gabor patch, orientation detectors could process the set in parallel and subsequently pool the information into a single percept. A similar conclusion was reached by Alvarez and Oliva (2009). An averaging system such as this is not directly dependent upon mechanisms of selective attention, because average orientation representation is believed to reflect an automatic, low-level physiological mechanism (Bosking, Crowley, & Fitzpatrick, 2002; Victor, Purpura, Katz, & Mao, 1994; Vogels, 1990).

Although it is clear that crowding (i.e., the inability to discriminate a target when it is flanked by distractors, even though it is perfectly discriminable when presented in isolation) is not necessary for the extraction of ensemble information, one intriguing possibility is that it enhances the precision of the summary representation. Chong and Treisman (2005a) showed that distributed attention improved average size representation; crowding (Evans & Chong, chapter 13 of this volume; Levi, 2008; Pelli, Palomares, & Majaj, 2004) by definition, disrupts any serial attentive process (Intriligator & Cavanagh, 2001), which may force observers into an attentional strategy more conducive to summary representation. Thus, perhaps crowding facilitates the condensation of (even consciously inaccessible) information into efficient "chunks."

PERCEIVING AVERAGE POSITION

Psychophysical experiments first demonstrated that humans are sensitive to average or centroid position (Hess & Holliday, 1992; Morgan & Glennerster, 1991; Whitaker, McGraw, Pacey, & Barrett, 1996). More recent work by Alvarez and Oliva (2008) suggests that selective attention may play a minimal role in this process. Using a multiple object tracking task (Intriligator & Cavanagh, 2001; Pylyshyn & Storm, 1988), Alvarez and Oliva (2008) found that, even when observers were unable to identify individual unattended objects, they could localize the centroid of those objects. Although Chong and Treisman (2005b) demonstrated that distributed attention could improve an estimate of the mean, this work (Alvarez & Oliva, 2008) showed that a summary might be derived beyond the focus of attention. Consistent with this, Demeyere and colleagues found that a patient with simultanagnosia (Balint syndrome, see Humphreys and Riddoch, chapter 15 of this volume; Robertson, chapter 14 of this volume) could perceive ensemble color in an array of stimuli despite being unaware of the array (Demeyere, Rzeskiewicz, Humphreys, & Humphreys, 2008).

PERCEIVING ENSEMBLES OF FACES

For many years, the focus of research on ensemble and summary statistical perception has been on low-level stimuli (motion, orientation, position, size, etc.). However, given our effortless interaction with highly complex scenes, and our subjective impression of a rich and complete visual world, it is reasonable to think that the ensemble coding heuristic might operate on a processing level beyond that of orientation, size, or texture. Haberman and Whitney (2007; 2009; Haberman, Harp, & Whitney, 2009) explored the possibility that observers could extract an average representation from high-level stimuli, including faces. The authors created a series of morphs, varying the expression of faces ranging from extremely happy to extremely sad. Observers viewed sets of these emotionally varying faces, and were asked whether a subsequent test face was happier or sadder than the mean expression of the previous set. Remarkably, observers could discriminate the average expression of the whole set as well as they could discriminate the expression of a single face. This phenomenon proved to be robust and flexible, operating implicitly and explicitly (Haberman & Whitney, 2009), across a variety of expressions as well as gender morphs (Haberman & Whitney, 2007), at short exposure durations (as low as 50 milliseconds, although with reduced precision; Haberman & Whitney, 2009) and on sets containing as many as 20 faces (Haberman, Harp, & Whitney, 2009; see figure 16.3 for a summary of results). Control experiments demonstrated that the mean discrimination of expression declined when viewing sets of inverted or scrambled faces, suggesting that the visual system extracts summary statistical information about the configural or holistic properties of faces, not just about low-level visual cues such as spatial frequency (Oliva & Torralba, 2001; Torralba & Oliva, 2003) or orientation. Summary statistical representation must occur at multiple, distinct levels of the visual processing hierarchy. High-level ensemble coding is further supported by other work showing that observers can rapidly perceive the mean identity of sets of faces (de Fockert & Wolfenstein, 2009), that the average expression is preserved even when the faces are crowded (Fischer & Whitney, 2011), and also by research showing rapid within-hemifield emotional averaging predicted by properties of neural averaging (Sweeny et al., 2009).

Perceived facial expression also rapidly integrates over time (Haberman, Harp, & Whitney, 2009). Observers viewed sequences of different faces presented at various temporal frequencies and made judgments about the mean expression of those sequences. The precision with which observers perceived average facial expression was relatively invariant to changes in temporal frequency. In fact, observers were able to accurately derive a mean expression in a sequence of 20 faces presented at 20 hertz. The more critical factor was the total time available for viewing the faces—curve fitting suggested that the time constant of temporal integration of perceived facial expression is around 800 milliseconds. Naturally, all visual processes require some amount of time to integrate (even motion and

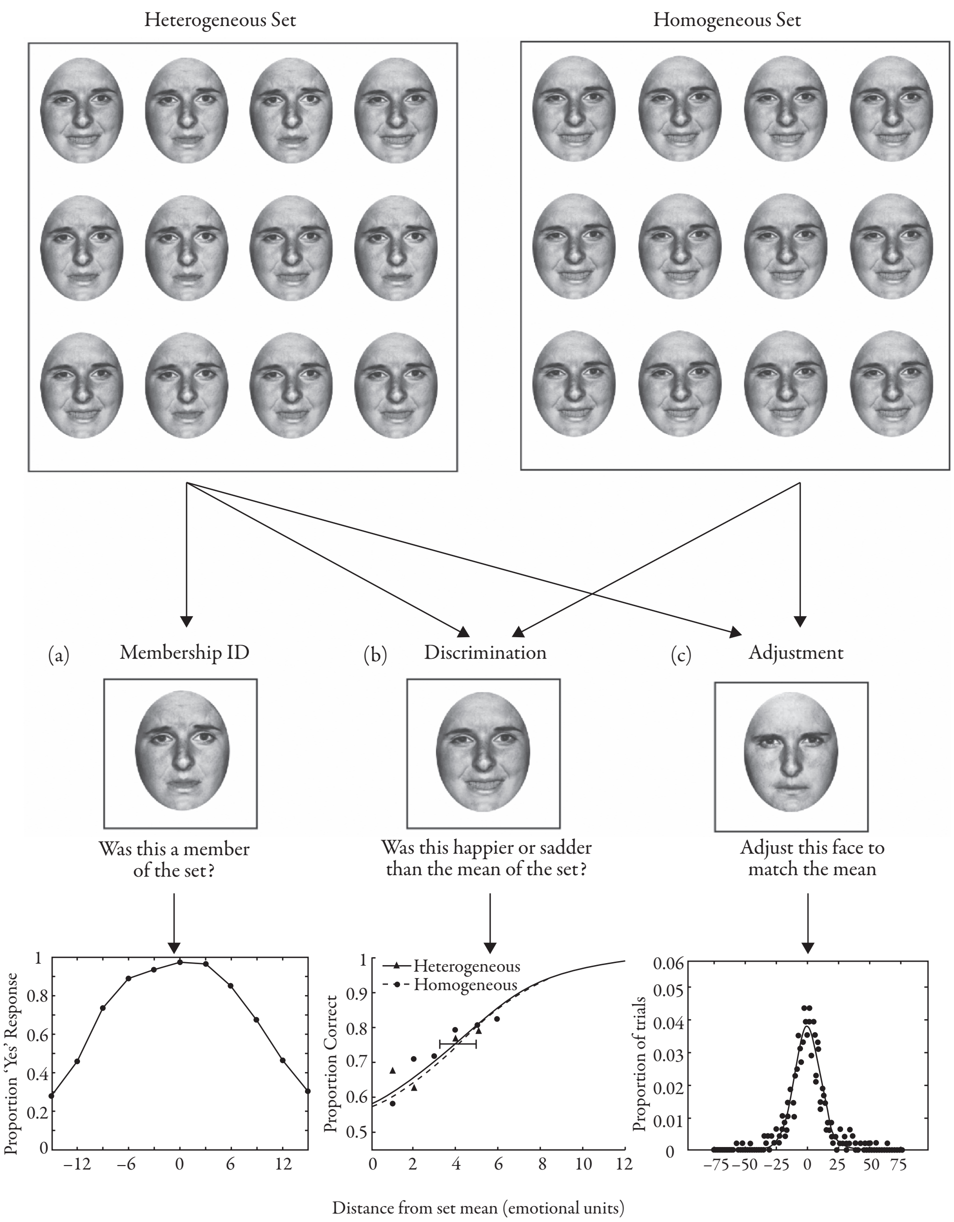

Figure 16.3 Some of the face-averaging paradigms implemented by Haberman and Whitney. a. Observers had to identify whether a test face was a member of the previously displayed set. Observers were most likely to indicate a test face was a set member when it approached the mean expression of the set (0 indicates the mean expression). Thus, observers were unable to represent the individual set constituents, but instead favored the ensemble. b. Observers were explicitly asked about the average expression in a set. Surprisingly, they could discriminate the mean expression (triangles, dotted line) as well as they could discriminate any single face (circles, solid line). **c.** Observers used the mouse to adjust the test face to match the mean expression of the set. This provided the full error distribution of the mean representation (0 indicates the mean expression). Responses tended to cluster around the mean expression of the set.

orientation, which are believed to be parallel processes). Therefore, an integration time constant of 800 milliseconds does not rule out the existence of a parallel mechanism at work. Although the integration time for sets of faces was higher than that for low-level motion (Burr, 1981; Nakayama, 1985; Snowden & Braddick, 1989), it compares favorably with the time it takes the visual system to perceive biological motion (Neri, Morrone, & Burr, 1998).

Summary statistics capture well and may explain texture appearance (the granite-ness, stucco-ness, etc.) of surfaces. Although textures have been extensively studied (Beck, 1983; Landy & Graham, 2004; Malik & Rosenholtz, 1997; Nothdurft, 1991), and summary statistical representation of low level features holds for typical "textures," the finding that groups of faces are perceived as an ensemble—as a texture—suggests that textures can occur at any level of visual analysis.

This brief survey is necessarily incomplete, but it provides a glimpse at the history of ensemble perception and some of the continuing debates. The next section explores some of these debates more fully, as well as some of the most common concerns regarding research on summary statistical perception.

CURRENT DEBATES IN SUMMARY STATISTICAL PERCEPTION

CAN ENSEMBLE PERCEPTION BE EXPLAINED BY SERIAL MECHANISMS OF ATTENTION?

One of the most important contributions of the work by Chong and Treisman (2003) is that it brought to light the possibility that summary statistical perception might provide an efficient means of representing information with little attentional involvement. Recently, the idea of an automatic averaging mechanism for absolute size has become more contentious. Some researchers have suggested that, unlike motion or orientation integration (Parkes et al., 2001; Watamaniuk & Duchon, 1992; Williams & Sekuler, 1984), the representation of average size is not supported from a physiological perspective (Myczek & Simons, 2008). Average motion and orientation perception may be driven by information pooled across individual receptors, whereas, according to Myczek and Simons (2008), there are no "size" receptors per se that can give rise to an average size percept (though cf. Op De Beeck & Vogels, 2000; Sripati & Olson, 2009; Stuart, Bossomaier, & Johnson, 1993; Vogels, 2009). Using a series of elegant simulations, Myczek and Simons (2008) argued that much of the extant average size data might be well captured by established mechanisms of selective attention.

The claim of average size automaticity was predicated on several pieces of evidence, including the invariant performance across set size manipulations (Ariely, 2001), implicit average size representation (Ariely, 2001), the speed with which average size was derived (accompanying paper), and the ability to represent the average size of subsets without a prior cue directing attention (Chong & Treisman, 2005b). Using existing average size discrimination datasets, Myczek and Simons (2008) modeled performance of an ideal observer who simply subsampled from the set of circles. That is, they examined mean discrimination performance when an observer examined N circles from a set of 12. These results have significant implications for the claims of a dedicated ensemble coding mechanism, at least in the arena of average size perception. If average- size discrimination could be explained by averaging a small portion of the set, it would suggest that well-established mechanisms of selective attention might be operating. Indeed, the simulations suggested that, for much of the published average size data, averaging just a couple of the items matched mean discrimination performance of actual observers (although some tasks required as many as four circles). Myczek and Simons (2008) noted that certain cognitive strategies (conscious or not) might serve to help observers "cheat" on the task. For example, in some of the paradigms, identifying the largest circle in each set was sufficient to correctly identify the set with the largest average size. However, Chong and Treisman (2003, accompanying paper; 2005b) used an extensive array of paradigms, manipulating the distribution of the circles, as well as the density and numerosity of the sets, which made a singular "cheat" strategy impossible. In a follow-up study, Chong and colleagues (Chong et al., 2008) had observers perform several variants of the average size discrimination paradigm within a single run. Observers were able to derive the average size equally well across the various paradigms, and the authors suggested that using a cognitive "cheat" would have required switching strategies on a trial-by-trial basis—something relatively improbable. In response to this demonstration, Simons and Myczek (Simons & Myczek, 2008) argued that, although empirical testing of alternative strategies was commendable, the manipulation did not discount the possibility of observers' utilizing still other strategies consistent with focused attention. They claimed that, because average-size perception depends on unknown physiological mechanisms (i.e., receptors sensitive to absolute size), the burden of proof rests on researchers to discredit subsampling, which operates under already established attentional mechanisms.

One shortcoming of the Myczek and Simons simulations, as pointed out in a recent commentary (Ariely, 2008), was that the authors did not incorporate any "judgment" noise. That is, the data represented an ideal observer. Estimating the proper amount of noise to incorporate is difficult at best, however, which is why empirically re-creating the conditions of the model using observers is prudent. Chong and colleagues (2008) asked observers to discriminate the average size given only a subset of the array of circles (one, two, or eight items), similar to the design used by Myczek and Simons in their simulations. The results suggested that estimating the average size when viewing one or two items could not match performance when observers viewed all eight items. This is in contrast to at least some of the modeling done by Myczek and Simons (2008), which showed that subsampling one or two items from the set was sufficient (in many cases) to match observer performance when viewing the whole set. However, Chong and colleagues (2008) did not characterize observer performance across additional sampling conditions (three through seven), which makes it difficult to assess just how much information was necessary to accurately represent the average size of the set. In addition, Simons contends that forcing observers to assess the average set size using arbitrarily assigned circles from the set may not reflect the strategy they use when given the opportunity to view the whole set (personal communication, 2009).

Work completed contemporaneously and independently may support the assertion that average size is computed via subsampling (de Fockert & Marchant, 2008). De Fockert and Marchant (2008) showed that when attention was directed to a specific constituent of the set (e.g., the largest circle in the display), the average size estimate was modulated by the size of the attended constituent. For example, observers tended to overestimate the average size of a set when they were instructed to attend to the largest item. Although this seems to implicate focused attention, it is actually only a modulation. Indeed, Chong and Treisman (2005a) have also demonstrated an attentional modulation of summary statistical perception. Attentional modulation, on its own, is not sufficient to adjudicate between the alternative explanations; that is, the fact that summary statistical perception is subject to attentional modulation does not imply that focused attention (i.e., subsampling) is the underlying strategy. Indeed, recent behavioral and modeling work in the face domain suggests that focused attention cannot account for average expression performance (Fischer & Whitney, 2011; Haberman & Whitney, 2010).

The issue of whether focused attention or parallel processes mediate average size perception remains an open question. Nevertheless, the recent line of inquiry highlights some critical ideas. For one, modeling behavior can be a powerful approach for ruling out specific alternative hypotheses, in this case the possibility that focused attention could be used as a strategy to derive the average size. Modeling, whenever possible, should be supported by empirical research to verify the model's plausibility. Although this modeling was used to test average size perception exclusively and says nothing of other summary statistical domains (Myczek & Simons, 2008; Simons & Myczek, 2008), it does raise the question of what constitutes a "parallel" process. Traditionally, the hallmark of a parallel process in the visual-search literature is a flat search slope as a function of the number of items in the display (Treisman & Gelade, 1980). In the case of averaging, set- size invariance has also been used to argue for a parallel mechanism. However, even processes like average size, speed, and orientation perception (Morgan et al., 2008; Parkes et al., 2001; Watamaniuk & Duchon, 1992; Watamaniuk et al., 1989; Williams & Sekuler, 1984), which are generally considered to operate in parallel, may also be explained by subsampling (e.g., Morgan et al., 2008). Therefore, what counts as a parallel process may not be parallel in the way that is often implied; it may not be that *every* item is analyzed simultaneously, or that every item is compared to every other item simultaneously, but, rather, a subset of items is analyzed as representative of the entire group (see figure 16.4 for a model of this idea). If the number of subsampled items required to match performance does not far exceed the limits of multiple object or attentive tracking (Pylyshyn & Storm, 1988), can we rule out attention? Conversely, even if the number of

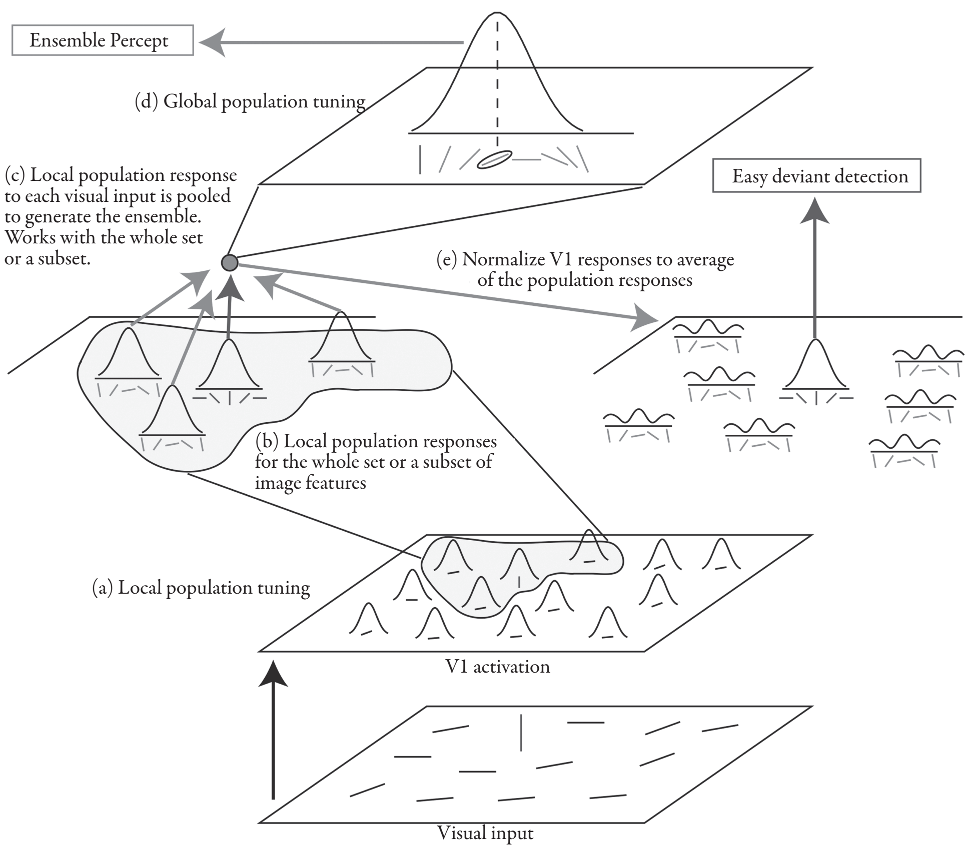

Figure 16.4 One possible physiological mechanism driving pop-out. a–b. Orientation selective cells (possibly in V1) fire in response to visual input. c–d. The activity from some or all of the orientation selective cells is combined to create the ensemble. e. Via feedback or horizontal connections, the activity from orientation selective cells is normalized to the population response (i.e., ensemble). Any cell activity remaining will correspond to the deviant. One of the strengths of this model is that it can operate in parallel, negating the computationally inefficient method of comparing each item with every other one.

sampled items is within attentional tracking and visual short term memory capacity limits (Luck & Vogel, 1997), then must we conclude that attention is responsible? Clearly not, as this test does not rule out a mechanism that simultaneously (truly in parallel) samples every item, but does so very noisily. As found in the temporal integration of faces (Haberman et al., 2009), there is a tradeoff between the number of samples and the noisiness of the sampling.

Other paradigms may lend themselves to disambiguating this issue more directly. For example, multiple object tracking, crowding, or global versus local orientation judgments (Alvarez & Oliva, 2008, 2009; Parkes et al., 2001) have been used to effectively demonstrate summary statistical representation beyond the focus of attention. However, these findings do not directly generalize to average size representation, ensuring continuing debate for some time to come.

IS ENSEMBLE PERCEPTION JUST A PROTOTYPE?

The demonstration of summary statistical representation for faces may raise the concern that the results are simply due to a prototype effect (Solso & Mccarthy, 1981). Indeed, there has been significant research providing evidence that observers implicitly develop statistical sensitivities to arbitrary patterns over time (Fiser & Aslin, 2001; Posner & Keele, 1968). However, unlike the prototype effect, ensemble coding requires no learning; summary statistical representation is a perceptual process and observers are sensitive to it after only a single trial. Prototype suggests that observers falsely recognize an average face due to predominant exposure to specific facial features over an extended period (Solso & Mccarthy, 1981). The average face (or size, orientation, etc.) in ensemble coding, though, changes on a trial-by-trial basis and is immediately recognizable. Ensemble perception is, therefore, a much more flexible pooling of important information into computationally palatable chunks. Observers never actually see the average face of a set and yet they favor the ensemble percept over the individuals.

MULTIPLE LEVELS AND MULTIPLE PATHWAYS OF ENSEMBLE CODING

The robust summary statistical representations found across domains suggest that ensembles are calculated at multiple levels in both the dorsal and ventral streams. Because orientation information is processed in early visual areas, average orientation is likely extracted prior to high-level object processing. Likewise, average expression from a crowd of faces must be mediated at a later stage of processing along the ventral pathway. Some ensembles, such as average brightness, color, and orientation, may be created at the earliest cortical (and possibly even subcortical) stages. Others, such as motion and position, may be generated along the dorsal stream. Finally, high-level shape and face ensembles are likely generated along the ventral, object-processing stream. Despite the distinct object properties processed at each level, the uniting commonality is that any set may be represented by a single, ensemble percept. This percept is created and maintained for conscious access, whereas the individual constituents are lost (via limitations of visual working memory, crowding, etc.). Because the visual system creates a representation of all the items within a set, loss of the individual is inconsequential. Many unanswered questions remain, such as how many concurrent ensemble percepts can be maintained, whether there is interference between different levels of ensemble analysis (e.g., average facial expression, brightness, and orientation), and whether the ensembles bypass the limited capacity of attention and visual short term memory, or instead simply act as "chunks" of information, increasing processing efficiency while still drawing on the finite resources of attention and memory.

IMPLICATIONS FOR ENSEMBLE CODING

VISUAL SEARCH

There is an appealing connection between ensemble coding and visual search. Despite the rich literature on the properties of visual search (Treisman, 1982; Verghese, 2001; Wolfe, Cave, & Franzel, 1989), a physiologically plausible mechanism (i.e., an algorithm and neural implementation in Marr's terms; Marr, 1982) driving pop-out (a phenomenon in which a visual target is rapidly discernable from a set of distractors) is still debated (Eckstein, 1998; Itti & Koch, 2000; Wolfe, 2003). Ensemble coding offers one possible solution: summary statistical representations may serve as a computationally efficient means of identifying deviance. Many models have made similar suggestions (e.g., Callaghan, 1984; Duncan & Humphreys, 1989). Usually, these models suggest that "similarity" modulates pop-out (Duncan & Humphreys, 1989). However, what counts as "similar" or "dissimilar" is unclear. Summary statistical representations per se could provide the underlying computation—the metric of similarity—that affords deviance detection.

How might such an algorithm that extracts ensemble information be implemented in the brain? Figure 16.4a shows an example of an array of oriented lines or Gabor patches, which would stimulate many local populations of orientation selective cells (e.g., in V1). If a subset (or the whole set) of *local* tuning curves is sampled (figure 16.4b) and pooled (figure 16.4c), a *global* population tuning curve is represented (the model would work equally well if the entire set were sampled, but this conception accounts for the possibility that only a subset of the items are sampled; c.f., [Dakin & Watt, 1997; Morgan et al., 2008; Myczek & Simons, 2008]. This global population curve is the average of local tuning curves and ultimately produces an ensemble percept (figure 16.4d). Note that the impact of any deviant orientation is washed out in the global population curve, because most of the inputs are of similar orientations. The global population response then normalizes the local tuning (via feedback or horizontal connections; figure 16.4e). Most of the local population responses are reduced to near 0, and what is left is activity corresponding to the deviant orientation. Although low-level normalization or contextually dependent procedures have been implemented in other models (e.g., Itti, Koch, & Niebur, 1998; Li, 1999), this model implicates ensemble coding as the physiological impetus for pop-out. A particular strength of this model is that the normalization operation may be carried out in parallel, without requiring multiple comparisons across local population responses.

LINKING SUMMARY STATISTICS TO PERCEPTION

There is ample evidence that humans are able to quickly extract a great deal of information from scenes (Oliva & Torralba, 2001; Potter, 1976; Thorpe, Fize, & Marlot, 1996; Torralba & Oliva, 2003). Exactly what cues reveal the gist of a scene and what particular physiological mechanism could code for this sort of information remain unknown. Although there are other possibilities, one intriguing idea is that it is summary statistics per se that drive much of what we consider or perceive as "gist" (e.g., Alvarez & Oliva, 2009; accompanying paper; Haberman & Whitney, 2007). There is already some support for this idea, but many open questions persist before a direct link can be formed between the seemingly incommensurable capacity limits of vision and the phenomenological richness of perception.

REFERENCES

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122–131.
Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science, 19*(4), 392–398.
Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Science U S A, 106*(18), 7345–7350.
Anderson, N. H. (1965). Averaging versus adding as a stimulus-combination rule in impression-formation. *Journal of Experimental Psychology, 70*(4), 394–400.
Anderson, N. H. (1971). Integration theory and attitude change. *Psychological Review, 78*(3), 171–206.
Anderson, N. H., Lindner, R., & Lopes, L. L. (1973). Integration theory applied to judgments of group attractiveness. *Journal of Personality and Social Psychology, 26*(3), 400–408.
Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*(2), 157–162.
Ariely, D. (2008). Better than average? When can we say that subsampling of items is better than statistical summary representations? *Perception & Psychophysics, 70*(7), 1325–1326.
Beck, J. (1983). Textural segmentation, 2nd-order statistics, and textural elements. *Biological Cybernetics, 48*(2), 125–130.
Bosking, W. H., Crowley, J. C., & Fitzpatrick, D. (2002). Spatial coding of position and orientation in primary visual cortex. *Nature Neuroscience, 5*(9), 874–882.
Britten, K. H., & Heuer, H. W. (1999). Spatial summation in the receptive fields of MT neurons. *Journal of Neuroscience, 19*(12), 5074–5084.
Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual-motion: A comparison of neuronal and psychophysical performance. *Journal of Neuroscience, 12*(12), 4745–4765.
Bulakowski, P. F., Koldewyn, K., & Whitney, D. (2007). Independent coding of object motion and position revealed by distinct contingent aftereffects. *Vision Research, 47*(6), 810–817.
Burr, D. C. (1981). Temporal summation of moving by the human visual-system. *Proceedings of the Royal Society of London Series B-Biological Sciences, 211*(1184), 321–339.
Burr, D., & Ross, J. (2008). A visual sense of number. *Current Biology, 18*(6), 425–428.
Callaghan, T. C. (1984). Dimensional interaction of hue and brightness in preattentive field segregation. *Perception & Psychophysics, 36*(1), 25–34.
Callaghan, T. C. (1989). Interference and dominance in texture segregation: Hue, geometric form, and line orientation. *Perception & Psychophysics, 46*(4), 299–311.
Chong, S. C., Joo, S. J., Emmanouil, T. A., & Treisman, A. (2008). Statistical processing: Not so implausible after all. *Perception & Psychophysics, 70*(7), 1327–1334.
Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*(4), 393–404.
Chong, S. C., & Treisman, A. (2005a). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics, 67*(1), 1–13.
Chong, S. C., & Treisman, A. (2005b). Statistical processing: Computing the average size in perceptual groups. *Vision Research, 45*(7), 891–900.
Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research, 37*(22), 3181–3192.
de Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception & Psychophysics, 70*(5), 789–794.
de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology (Colchester)*, 1–7.
Demeyere, N., Rzeskiewicz, A., Humphreys, K. A., & Humphreys, G. W. (2008). Automatic statistical processing of visual properties in simultanagnosia. *Neuropsychologia, 46*(11), 2861–2864.
Duncan, J., & Humphreys, G. W. (1989). Visual-search and stimulus similarity. *Psychological Review, 96*(3), 433–458.
Eckstein, M. P. (1998). The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. *Psychological Science, 9*(2), 111–118.
Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12*(6), 499–504.
Fischer, J. & Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology, 106*(3), 1389–1398.
Frechette, E. S., Sher, A., Grivich, M. I., Petrusca, D., Litke, A. M., & Chichilnisky, E. J. (2005). Fidelity of the ensemble code for visual motion in primate retina. *Journal of Neurophysiology, 94*(1), 119–135.
Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision, 9*(11), 1–13.
Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology, 17*(17), R751-R753.
Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance, 35*(3), 718–734.
Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention Perception & Psychophysics, 72*(7), 1825–1838.
Hess, R. F., & Holliday, I. E. (1992). The coding of spatial position by the human visual-system: Effects of spatial scale and contrast. *Vision Research, 32*(6), 1085–1097.
Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43*(3), 171–216.
Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10–12), 1489–1506.
Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(11), 1254–1259.
Jancke, D. (2000). Orientation formed by a spot's trajectory: A two-dimensional population approach in primary visual cortex. *Journal of Neuroscience, 20*(14), RC86.
Koenderink, J. J., van Doorn, A. J., & Pont, S. C. (2004). Light direction from shad(ow)ed random Gaussian surfaces. *Perception, 33*(12), 1405–1420.

Koffka, K. (1935). *The principles of gestalt psychology*. London: Routledge and Kegan Paul.
Landy, M., & Graham, N. (2004). Visual perception of texture. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (Vol. 2, pp. 1106–1118). Cambridge, MA: MIT Press.
Leon, M., Oden, G. C., & Anderson, N. H. (1973). Functional measurement of social values. *Journal of Personality and Social Psychology, 27*(3), 301–310.
Levi, D. M. (2008). Crowding - An essential bottleneck for object recognition: A mini-review. *Vision Research, 48*(5), 635–654.
Levin, I. P. (1974). Averaging processes in ratings and choices based on numerical information. *Memory & Cognition, 2*(4), 786–790.
Li, Z. P. (1999). Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proceedings of the National Academy of Sciences of the United States of America, 96*(18), 10530–10535.
Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279–281.
Malik, J., & Rosenholtz, R. (1997). Computing local surface orientation and shape from texture for curved surfaces. *International Journal of Computer Vision, 23*(2), 149–168.
Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: W.H. Freeman.
Morgan, M., Chubb, C., & Solomon, J. A. (2008). A 'dipper' function for texture discrimination based on orientation variance. *Journal of Vision, 8*(11), 1–8.
Morgan, M. J., & Glennerster, A. (1991). Efficiency of locating centres of dot-clusters by human observers. *Vision Research, 31*(12), 2075–2083.
Morgan, M. J., Watamaniuk, S. N. J., & McKee, S. P. (2000). The use of an implicit standard for measuring discrimination thresholds. *Vision Research, 40*(17), 2341–2349.
Motoyoshi, I., & Nishida, S. (2001). Temporal resolution of orientation-based texture segregation. *Vision Research, 41*(16), 2089–2105.
Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics, 70*(5), 772–788.
Nakayama, K. (1985). Biological image motion processing - A review. *Vision Research, 25*(5), 625–660.
Neri, P., Morrone, M. C., & Burr, D. C. (1998). Seeing biological motion. *Nature, 395*(6705), 894–896.
Newsome, W. T., & Pare, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (Mt). *Journal of Neuroscience, 8*(6), 2201–2211.
Noe, A., Pessoa, L., & Thompson, E. (2000). Beyond the grand illusion: What change blindness really teaches us about vision. *Visual Cognition, 7*(1–3), 93–106.
Nothdurft, H. C. (1991). Texture segmentation and pop-out from orientation contrast. *Vision Research, 31*(6), 1073–1078.
Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42*(3), 145–175.
Op De Beeck, H., & Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. *Journal of Comparative Neurology, 426*(4), 505–518.
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.
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.
Posner, M. I., & Keele, S. W. (1968). On genesis of abstract ideas. *Journal of Experimental Psychology, 77*(3), 353–363.
Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology-Human Learning and Memory, 2*(5), 509–522.
Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3*(3), 179–197.
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 & Performance, 18*(3), 849–860.
Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*(5), 368–373.
Sanders, K., Haberman, J., & Whitney, D. (2008). Mean representation beyond a shadow of a doubt: Summary statistical representation of shadows and lighting direction. *Journal of Vision, 8*(6), 334–334a.
Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology, 38*(2), 259–290.
Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception, 28*(9), 1059–1074.
Simons, D. J., & Myczek, K. (2008). Average size perception and the allure of a new mechanism. *Perception & Psychophysics, 70*(7), 1335–1336.
Snowden, R. J., & Braddick, O. J. (1989). The combination of motion signals over time. *Vision Research, 29*(11), 1621–1630.
Solso, R. L., & Mccarthy, J. E. (1981). Prototype formation of faces - A case of pseudo-memory. *British Journal of Psychology, 72*, 499–503.
Sripati, A. P., & Olson, C. R. (2009). Representing the forest before the trees: A global advantage effect in monkey inferotemporal cortex. *Journal of Neuroscience, 29*(24), 7788–7796.
Stuart, G. W., Bossomaier, T. R. J., & Johnson, S. (1993). Preattentive processing of object size—Implications for theories of size perception. *Perception, 22*(10), 1175–1193.
Sweeny, T. D., Grabowecky, M., Paller, K., & Suzuki, S. (2009). Within-hemifield perceptual averaging of facial expression s predicted by neural averaging. *Journal of Vision, 9*(3), 1–11.
Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520–522.
Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network-Computation in Neural Systems, 14*(3), 391–412.
Treisman, A. (1982). Perceptual grouping and attention in visual-search for features and for objects. *Journal of Experimental Psychology-Human Perception and Performance, 8*(2), 194–214.
Treisman, A., & Gelade, G. (1980). Feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97–136.
Treisman, A., & Gormican, S. (1988). Feature analysis in early vision - Evidence from search asymmetries. *Psychological Review, 95*(1), 15–48.
Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron, 31*(4), 523–535.
Victor, J. D., Purpura, K., Katz, E., & Mao, B. Q. (1994). Population encoding of spatial-frequency, orientation, and color in macaque V1. *Journal of Neurophysiology, 72*(5), 2151–2166.
Vogels, R. (1990). Population coding of stimulus orientation by striate cortical-cells. *Biological Cybernetics, 64*(1), 25–31.
Vogels, R. (2009). Visual perception: Larger is faster. *Current Biology, 19*(16), R691-R693.
Watamaniuk, S. N. J., & Duchon, A. (1992). The human visual-system averages speed information. *Vision Research, 32*(5), 931–941.
Watamaniuk, S. N. J., & McKee, S. P. (1998). Simultaneous encoding of direction at a local and global scale. *Perception & Psychophysics, 60*(2), 191–200.
Watamaniuk, S. N. J., Sekuler, R., & Williams, D. W. (1989). Direction perception in complex dynamic displays - The integration of direction information. *Vision Research, 29*(1), 47–59.
Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. *Psychologische Forschung, 4*, 301–350.
Whitaker, D., McGraw, P. V., Pacey, I., & Barrett, B. T. (1996). Centroid analysis predicts visual localization of first- and second-order stimuli. *Vision Research, 36*(18), 2957–2970.
Williams, D. W., & Sekuler, R. (1984). Coherent global motion percepts from stochastic local motions. *Vision Research, 24*(1), 55–62.
Wolfe, J. M. (2003). Moving towards solutions to some enduring controversies in visual search. *Trends in Cognitive Sciences, 7*(2), 70–76.
Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search - An alternative to the feature integration model for visual-search. *Journal of Experimental Psychology-Human Perception and Performance, 15*(3), 419–433.