CellPress

Opinion

What is the Bandwidth of Perceptual Experience?

Michael A. Cohen,[1,*] Daniel C. Dennett,[2] and Nancy Kanwisher[1]

Although our subjective impression is of a richly detailed visual world, numerous empirical results suggest that the amount of visual information observers can perceive and remember at any given moment is limited. How can our subjective impressions be reconciled with these objective observations? Here, we answer this question by arguing that, although we see more than the handful of objects, claimed by prominent models of visual attention and working memory, we still see far less than we think we do. Taken together, we argue that these considerations resolve the apparent conflict between our subjective impressions and empirical data on visual capacity, while also illuminating the nature of the representations underlying perceptual experience.

Perception: Rich or Sparse?

The moment we open our eyes, we experiences a vast, richly detailed visual world extending well into the periphery [1,2]. However, numerous experimental results indicate that the bandwidth of human perception is severely limited. Findings from change blindness and inattentional blindness demonstrate that much of the available visual information goes unnoticed [3]. Direct estimates of the capacity of visual **attention** (see Glossary) and working memory reveal that surprisingly few items can be processed and maintained at once [4,5]. These results raise a natural question: why do we think we see so much when the scientific evidence suggests we see so little?

One answer to this question is that change blindness and inattentional blindness highlight the limits of mechanisms such as attention and working memory, rather than the limits of conscious perception. According to this view, perception 'overflows' and exceeds the capacity of the cognitive mechanisms needed to access that information [6]. In other words, we consciously perceive more than we can attend, remember, report, or base decisions on [7–11]. Under this view, the neural processes associated with visual **awareness** are separate from those associated with attention, working memory, and explicit report. **Recurrent processing** in sensory cortex supports conscious perception [10], whereas the parietal and prefrontal cortices support the cognitive mechanisms involved in accessing those percepts [12]. According to this framework, there is no tension between our subjective impression of the world and objective measures of human capacity limits because both of these are true. We have a rich experience of the world that cannot be fully captured by the capacity-limited cognitive mechanisms beyond the canonical visual system.

However, contrary to this view, many researchers argue that awareness is intrinsically linked to these cognitive functions and information is not consciously perceived until it is accessed by higher-order systems, such as attention, working memory, and decision-making [13–18]. Rather than link conscious perception with recurrent processing in sensory cortex, this view associates awareness with the parietal and prefrontal cortices [14]. However, for those who endorse this view, the problem remains: how can our impression of a rich visual experience be supported by mechanisms that have strict capacity limits? Put another way, it has been claimed

Trends

Numerous empirical results highlight the limits of visual perception, attention, and working memory. However, it intuitively feels as though we have a rich perceptual experience, leading many to claim that conscious perception overflows these limited cognitive mechanisms.

A relatively new field of study (visual ensembles and summary statistics) provides empirical support for the notion that perception is not limited and that observers have access to information across the entire visual world.

Ensemble statistics, and scene processing in general, also appear to be supported by neural structures that are distinct from those supporting object perception. These distinct mechanisms can work partially in parallel, providing observers with a broad perceptual experience.

Moreover, new demonstrations show that perception is not as rich as is intuitively believed. Thus, ensemble statistics appear to capture the entirety of perceptual experience.

[1]Department of Brain and Cognitive Sciences, McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, USA
[2]Center for Cognitive Studies, Department of Philosophy, Tufts University, Medford, MA, USA

*Correspondence:
michaelthecohen@gmail.com
(M.A. Cohen).

CrossMark

that 'Introspectively, consciousness seems rich in content. . .From the third-person perspective of the behavioral scientist, however, consciousness is rather miserable' ([10] p. 205).

We argue here that, even though conscious perception is limited by cognitive mechanisms such as attention and working memory [3], it is not 'rather miserable', and the visual information observers have access to is not at all sparse. To make this argument, we discuss a variety of recent results demonstrating that people can encode and remember considerably more than just a few items. First, we examine empirical findings from a relatively new field of study: visual **ensembles and summary statistics** [19]. The key idea here is that the visual system exploits the redundancy found in real-world scenes to represent a large amount of information, often extending into the visual periphery, as a single summary statistic [20]. Critically, standard models of attention and working memory largely ignore ensemble representations, focusing instead on the representation of individual items [21–25]. Once ensembles and summary statistics are taken into consideration, it quickly becomes clear that observers have access to different aspects of the entire field of view, not just a handful of items.

In addition, we also discuss the idea that neural structures within the visual system involved in representing visual scenes and ensemble statistics [26,27] comprise a unique neural channel that is partially separate from other processing channels [28,29]. These results suggest that the visual system is functionally organized to allow for scene and ensemble representations to be efficiently formed somewhat independently of other object representations. In other words, there appear to be separate neural pathways for representing the forest and the trees.

Together, these findings help reconcile the apparent tension between our subjective impression of a rich visual world and empirical results highlighting the limits of visual cognition. We argue that the apparent richness of visual experience can be captured without having to dissociate consciousness from higher-level cognitive functions and without arguing that visual awareness overflows cognitive access.

The Limits of Visual Cognition

Two paradigms that have had a major role in demonstrating the limits of visual cognition are change blindness and inattentional blindness. Change blindness is the inability to detect a change between two different pictures when a brief interruption occurs between the two images [30,31] or the change occurs so gradually that it does not automatically draw attention [32]. By contrast, inattentional blindness is the failure to notice an otherwise visible stimulus when attention is directed elsewhere. In perhaps the most famous example, participants failed to notice a man in a gorilla costume walking through the middle of a scene when attention was focused on people passing a basketball [33]. Perhaps more commonly, automobile accidents regularly occur because drivers fail to notice items on the road (e.g., another car or a pedestrian) when their attention is directed elsewhere (e.g., their cell phone conversation) [34,35]. Despite differences in methodologies, both change blindness and inattentional blindness arise because of observers' limited ability to attend to and remember more than a few items at a time.

Although these paradigms clearly demonstrate the limits of visual cognition, more targeted studies have characterized the architecture and capacities of visual attention and working memory. Both of these processes are limited by a finite supply of some mental commodity [36]. This commodity is often characterized as either a fixed number of 'slots' [4,22–24] or a fluid cognitive resource [21,37,38]. Despite the differences between these models, they both converge on the idea that observers can store around three or four items in working memory. In terms of visual attention, initial studies estimated that around three or four locations can be attended at once [39,40], but more recent efforts have pushed that number closer to around seven or eight [25,41]. However, even eight attended locations is still not sufficient to explain the richness of perception.

Glossary

Attention: the process of selecting some bits of information for further processing at the expense of others (e.g., attending to the sound of a lecturer's voice and ignoring the street noise outside).
Awareness: the ability to consciously perceive, feel, or experience certain sensory events.
**Bayesian inference:** a method of statistical inference that uses Bayes' theorem to update the probability for a hypothesis as more information and/or evidence becomes available.
Ensembles and summary statistics: the representation of multiple items in the world as a single, average descriptor of the whole set (e.g., the average size of a collection of objects).
Gist of the scene: the basic perceptual (i.e., color, etc.) and conceptual (i.e., semantic label, etc.) representations of a scene that observers can comprehend in a single glance.
Recurrent processing: corticocortical interactions between neural regions in which information is transmitted from higher-level regions back to lower-level regions (e.g., from higher-level cortex back to early visual cortex).
Saccades: quick movements of the eyes that change the point of fixation.
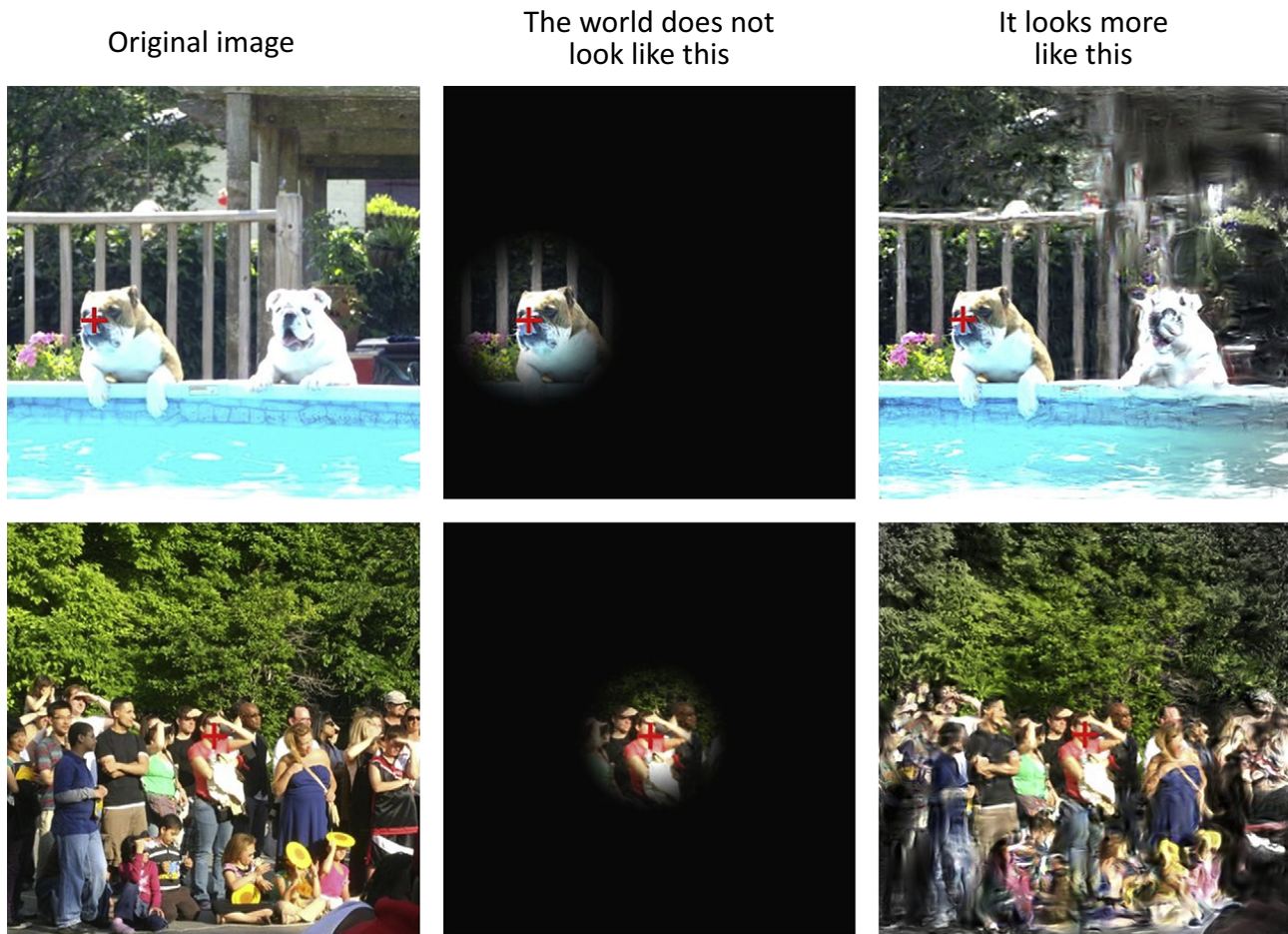
Original image	The world does not look like this	It looks more like this

Figure 1. An Approximate Visualization of a Snapshot of Visual Perception. The red crosses represent the fixation spot and the focus of attention for these particular snapshots. The left column shows two images that have not been doctored in any way (left column, 'Original image'). The middle column shows how the world does not fade into darkness when it is not fixated or focally attended. Instead, the right column shows how the parts of the world that are not fixated and unattended are represented as an ensemble statistic. These images were initially created using a texture synthesis model that preserves a variety of low-level statistics within a series of pooling regions similar to those of receptive fields [100]. In addition, these images are also shown through a circular Gaussian aperture so that the coloring of the images gradually disappears towards the periphery.

In isolation, these results seem to imply that awareness is limited to only a handful of items at a given moment. However, even when attention is entirely focused on a single item, no one has the impression that the rest of the world fades into darkness (Figure 1). Instead, observers believe they have a rich perceptual experience that spans the entire field of view. This belief has been experimentally verified by the fact that naïve observers systematically overestimate the capacities of attention and working memory [42,43]. At first blush, these results challenge the idea that the contents of visual awareness are the same as the contents of mechanisms such as attention and working memory [13–18]. How can such limited processes ever capture the richness of perceptual experience?

Ensemble Statistics and the Capacity of Perception

The visual world does not comprise random bits of uncorrelated information; it has structure, regularity, and redundancy [44–46]. The visual system takes advantage of this fact by representing groups of items as a statistic that summarizes different types of information (Box 1).

> ### Box 1. Ensembles and Summary Statistics
>
> What kind of information can be represented as an ensemble? Earlier studies focused on low-level visual dimensions, such as average orientation [101], brightness [102], speed of motion [103], and size [104]. These findings were then extended into higher-level dimensions, such as facial emotion, gender [105], and identity [106], as well as eye gaze [107] and biological motion [108]. Many of these dimensions can be processed remarkably fast (i.e., with ∼50 ms presentation time) [104,109] and formed by integrating representations over time [110,111], providing observers with a rapidly updated summary of a variety of dimensions across the visual world.
>
> Ensemble perception is not limited to the laboratory and is pervasive throughout everyday life. Imagine walking down a busy street with a crowd of people moving towards you. It would be computationally taxing to examine every object on the street or proceed person by person to determine each individual's facial expression, direction of motion, and gait. Given the inherent structure in the scene, a vast amount of this information spanning a wide expanse of visual space can be represented as an ensemble, or average. Representing information in this way allows observers to quickly determine whether the people are approaching in a threatening manner (e.g., moving quickly, in a converging direction, with angry facial expressions) or a nonthreatening manner (e.g., moving slower, in multiple directions, with neutral facial expressions). This is just one of many examples of how ensemble perception can enable efficient coding of relevant information as observers navigate the world despite processing limitations.

These ensemble representations, or summary statistics, are formed by collapsing across the measurements of individual items to form a singular description of the group. Although items that are not focally attended are represented with poor resolution, averaging across these imprecise representations allows the system to obtain an accurate measure of the entire group [20].

How does representing multiple items as an ensemble help resolve the tension between our subjective impression of a rich visual experience versus objective measurements of limited perception? We argue that items that are attended to and foveated are perceived at a higher resolution, while items that unattended or are in the periphery are primarily perceived as being part of an ensemble [47]. Observers are aware not only of a handful of items but also of the entire scene, but they only perceive a subset of the scene at high resolution. Standard demonstrations of change blindness and inattentional blindness succeed because the critical change often preserves the summary statistics of the scene. When those statistics are violated, it is considerably easier to notice the changes in a scene [48–50]. Detecting changes in these statistics is easy because observers do not perceive a small subset of the items in a scene; they perceive some information about all of the items in the form of ensemble statistics across several dimensions (Figure 1).

One of the cleanest demonstrations of this idea comes from a study in which participants performed a change detection task in which one of 25 colored items could change color (e.g., from red to blue) and participants simply indicated whether change occurred [51]. Performance was measured as a function of the statistical regularity of the display, which varied across trials. If observers can attend to, perceive, and remember only a handful of items (around three or four), performance should remain constant regardless of the changes to the structural configurations of the display. However, if observers are able to perceive the overall structure of the display, performance on the task should vary as a function of higher-order regularities. The results from this study unambiguously supported the latter prediction. When the display had little structure, standard estimates of working memory capacity [52] indicated that only around 4.5 items were successfully held in memory. However, when more structure was added to the display, an estimated 24 items were held in memory, six times the standard estimate of working memory capacity (Figure 2, top row). Furthermore, the authors modeled this change detection task using **Bayesian inference**. The results of this modeling suggest that observers encoded a few individual items along with a summary of the statistics of the entire display (see also [53–55]).

Ensemble Statistics and Natural Scenes

Ensemble statistics are useful not only for the simple stimulus displays typical of working memory paradigms, but also because probably they serve as the foundation of scene perception more

Perception is sparse:
Hard to spot the difference

Perception is rich:
Easy to spot the difference
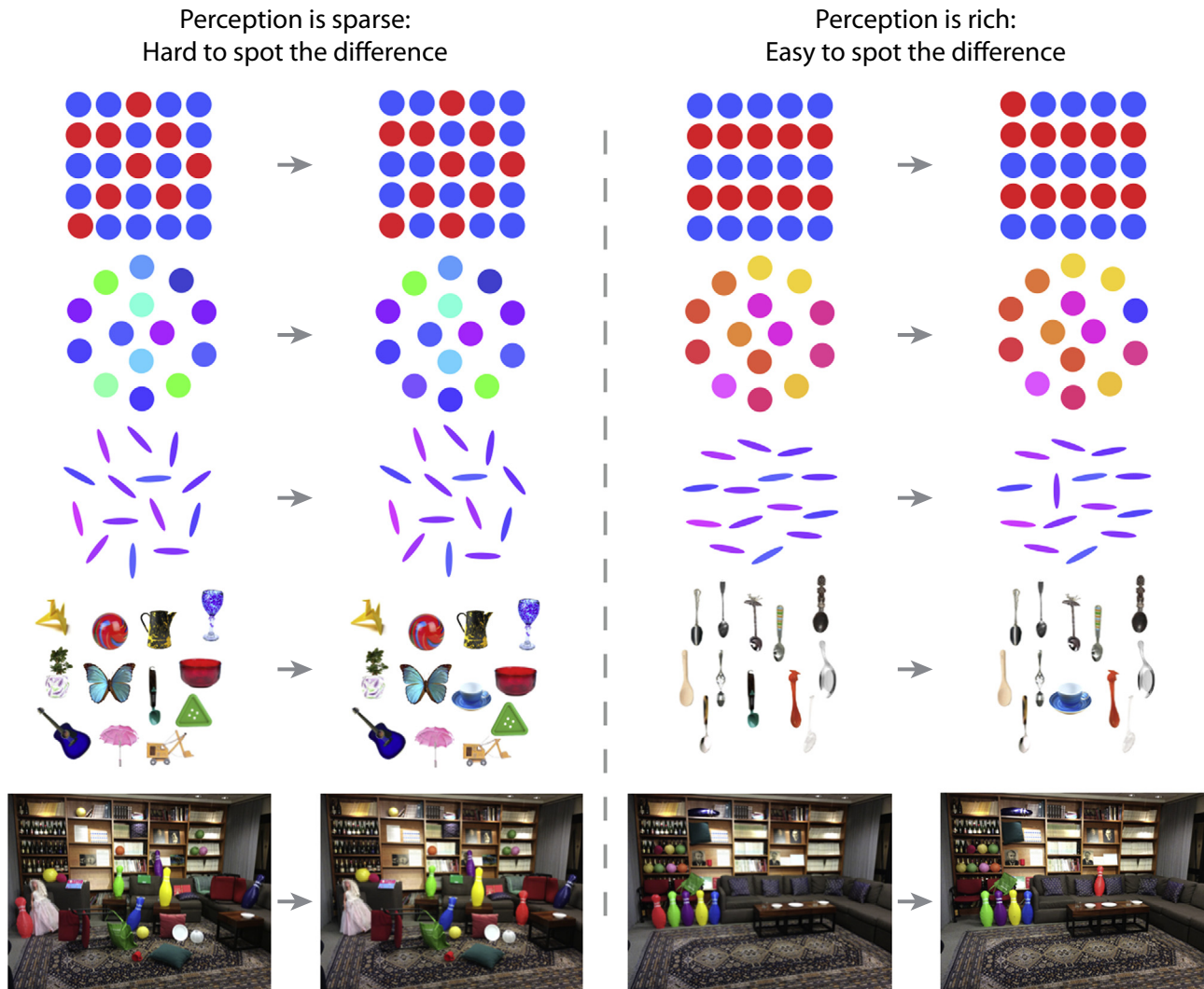
Trends in Cognitive Sciences

Figure 2. Change Blindness Demonstrations Are often Cited as Evidence that Perception Is Limited to just a few Items. The left two columns show five examples of displays in which detecting the change between them is difficult. However, the right two columns show five examples of displays in which detecting the change is trivially easy. Critically, the amount of change on each row is identical in both the left and right image pairs. The changes on each row are: (i) a red dot changes to a blue dot; (ii) a color changes by 90° on the color wheel; (iii) a blue oval rotates 90°; (iv) a spoon changes into a tea cup; and (v) the red bowling pin changes location. The key point is that, when the changes between the displays preserve certain statistics (left columns), the change is hard to notice. However, when the changes violate those statistics (right column), the change is easy to notice and highlights the vast amount of information observers have access to at a given instant.

broadly. Low-level features, such as luminance, orientation, and spatial frequency, are combined to form higher-order representations [56] that are sufficient for the classification of scenes (e.g., mountain, highway, or beach) [57]. As long as certain statistics of the scene are preserved (e.g., spatial contours, texture densities, etc.), the category of a scene can be extracted even when the individual objects it contains can no longer be perceived [58] (Figure 3). The importance of summary statistics has also been demonstrated with computational models that categorize scenes based solely on texture statistics [59]. Finally, in addition to carrying information about the identity of a scene, these basic statistics are informative enough for observers to recognize other aspects of a scene, such as its openness, symmetry, complexity, and depth [60,61].

| Original image | Preserved statistics | Original image | Preserved statistics |

Trends in Cognitive Sciences

Figure 3. Visualizations of How Many Aspects of a Scene Can Be Conveyed by only a Few Basic Image Statistics. Each of these images was generated by taking the original images and preserving certain global statistics in nonoverlapping windows [112].

Ensemble statistics also likely have a key role in enabling observers to form scene representations extremely quickly. These representations are formed so fast that observers can perceive a great deal of information from a single eye fixation, without making **saccades** [62]. When looking at a scene, fixations typically last 275–300 ms [63,64]. In that time, observers can extract the **gist of the scene** [65] and a few larger objects in the scene [66]. Even with an exposure duration of 50–100 ms, observers can still report the gist of a scene and extract a variety of properties, such as the depth, navigability, openness, and the temperature of the scene [67–70].

Together, these studies show that, within a single glance, observers do not merely have access to a small handful of isolated items in a sea of nothingness; they have access to a tremendous amount of information spanning the entire scene. The ability to extract an almost immediate sense of the visual world provides ecological benefits, such as guiding further action, especially saccades. Saccades are important to this discussion because one reason why observers may overestimate their perceptual experience is that saccades are so effortless that observers often do not even realize that they are making them [71]. This gives people the false impression that they perceive more than they actually do in a given instant because they are not aware of the serial manner in which they accrue information (i.e., one saccade after another). Furthermore, observers do not move their eyes randomly across a scene; they systematically go to the parts of the scene that are most informative for the task at hand [72]. This ability to select saccade targets intelligently is possible because observers are able to take advantage of the knowledge they have obtained about the scene from its global image statistics [73–75]. Thus, the use of summary statistics in scene perception not only gives observers an immediate sense of the visual world, but also provides a foundation for further exploration.

Finally, we speculate that, in addition to recognizing basic perceptual aspects of the world in a single glance, observers can also quickly perceive higher-level aspects of the scene, such as its physical, social, and action-based properties. For example, we do not see just a cup and a table, but a cup resting on a table, and we compute the reaching and grasping motion that would be required to take a sip from the cup. When seeing people in a scene, we quickly perceive their social characteristics and actions (e.g., are these two people interacting with each other or not?). If these inferences are made efficiently, it may due to specialized cortical machinery that helps imbue real-world perception with rich semantic content far beyond the mere identities of objects and scenes. Whether and how these high-level inferences bypass the standard processing bottlenecks of vision is an important and largely unanswered question for future research (see Outstanding Questions).

Neural Mechanisms of Scene and Ensemble Perception

While the speed and efficiency of natural scene and ensemble perception is well known and incorporated into many cognitive models [76,77], the neural mechanisms supporting this ability have only begun to be understood. A series of recent studies all converge on the idea that the gist and statistics of a scene are perceived so efficiently because we have neural structures specifically involved in representing those particular visual dimensions [78]. The parahippocampal place area (PPA), restrosplenical cortex (RSC), and occipital place area (OPA) are selectively and causally involved in recognizing the identity, layout, and navigability of scenes [79–81]. In addition, some of these neural regions, particularly PPA, appear to have a prominent role in representing a variety of ensemble and statistical properties (e.g., texture) [27,82,83].

The neural structures that are sensitive to scenes and ensemble statistics appear to be at least partially separate from the structures involved in representing other visual categories, such as faces, bodies, and objects. For example, neuropsychological studies have shown perserved texture and scene perception, but impaired object perception, in a patient with bilateral damage to lateral occipital cortex, an area of the brain that responds selectively to shape and object stimuli [84,85]. Conversely, a patient with damage to parahippocampal cortex had impaired scene recognition abilities and could only recognize scenes because of a prominent visual object (e.g., a house) [86]. Furthermore, behavioral and neural evidence from normal observers suggests that both scene and ensemble perception draw upon pools of cognitive resources that are distinct from pools supporting object perception [28,29,87]. One recent study found that more information could be processed when it was distributed across multiple neural regions (e.g., two faces and two scenes) compared with when it relied on a single neural region (e.g., four faces) [28] (Figure S1 in the supplemental information online). Thus, the visual system appears to be organized so that the representations of scenes, and potentially the representations of ensemble statistics, are formed with minimal interference from other objects. Having dedicated neural structures for these particular visual dimensions potentially has an important role in the ability to construct a richly detailed percept from a single glance.

Do Observers Have Access to all of this Information?

Those who believe that visual awareness overflows the capacities of attention, working memory, and other higher-level cognitive processes [7–11] may claim that the ensemble statistics described here are in fact prominent examples of such overflow. In fact, a recent study claimed that one particular statistic, color diversity, could be perceived 'cost free' and required no attention or working-memory resources [8]. Are ensemble statistics and scene representations truly 'cost free'? If this is true, ensembles and scenes should be immune to all types of attentional interference. Performing a demanding attentional task (e.g., visual search or working memory) should have no impact on the ability to perceive ensemble statistics or a natural scene.

In reality, numerous pieces of evidence suggest that some type of attention is needed to process and perceive ensembles and scenes. It has repeatedly been shown that natural scenes can go unnoticed because of inattentional blindness [3] or the attentional blink [88]. Furthermore, the ability to classify the gist of a scene suffers in dual-task situations [89]. In terms of ensemble statistics, one recent study found that processing the statistics of a group of objects requires as much attention as processing an individual object [90]. This finding is consistent with earlier studies claiming that an ensemble takes up approximately the same amount of space in working memory as an individual object [91–93]. In addition, observers are more accurate at processing multiple ensembles when they are presented sequentially, rather than simultaneously [94,95], suggesting that ensembles compete for limited cognitive resources. Finally, the precision of ensemble representations varies as a function of the allocation of attention (i.e., focused versus distribute) [96]. Together, these results suggest that, although ensemble statistics are processed quickly and efficiently [47], they are not perceived 'cost free.' However, this is a relatively new

---

Box 2. How to Demonstrate that Perception Is Not as Rich as it Seems

First, ask a participant how close a playing card has to be to the center of their field of vision for them to determine the identity of the card (e.g., ten of clubs versus seven of spades). Have the participant hold his arms up to visualize an approximate estimate of his answer (Figure IA). Next, to show the participant the actual answer, tell him to extend his arm to the side (~90° from fixation). Put a card in his hand facing him, but make sure that he keeps looking forward and does not glance at the card. Then, tell him to keep the card at arm's length and slowly move it in to the center of his field of view. Tell him to stop moving the card as soon as he is sure he can identify the card. If done correctly, it will become clear on the first trial that people wildly underestimate how close the card has to be to fixation for them to identify it. In many cases, participants will spontaneously laugh once they realize how far off they were with their prediction.

For the second demonstration, ask a participant to hold his fixation on a random object. Grab a colored object that fits in your hand (e.g., an orange). Slowly move your hand from the participant's periphery to the center of his field of view (Figure IB). Make sure to emphasize that he keep staring straight ahead. As you move your hand towards the center, jiggle it a little bit. Tell the participant to say 'Stop' as soon as he detects any motion in their periphery. Really emphasize that, as soon as he senses any peripheral motion, he should tell you immediately. Once he says stop, jiggle the object for a moment longer and confirm that he sees peripheral motion. If you want to confirm that he sees something, move your hand up/down and ask the participant to say which direction your hand moved. Once you are both convinced he can see the peripheral motion, ask him to tell you the color of the object. He will almost certainly say he does not know and if you force him to give you an answer, he will just guess. Do this a multiple times and you will see that people are: (i) no better than chance at guessing the color of the object; and (ii) surprised by how unable they are to report the color.
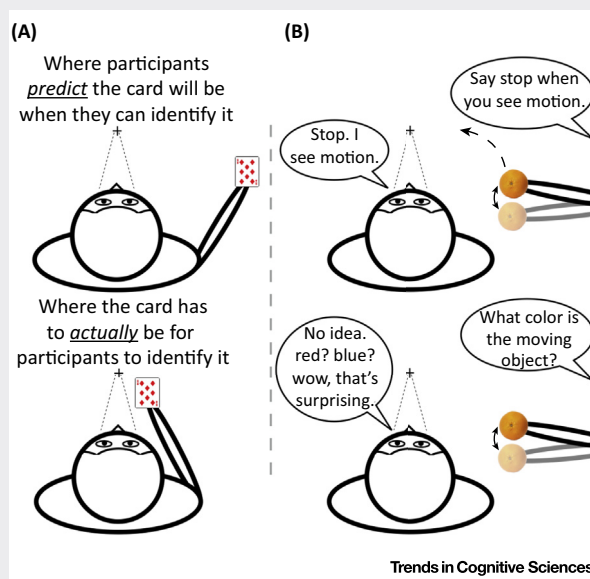

(A) Where participants _predict_ the card will be when they can identify it

Where the card has to _actually_ be for participants to identify it

(B) Say stop when you see motion.

Stop. I see motion.

No idea. red? blue? wow, that's surprising.

What color is the moving object?

Trends in Cognitive Sciences

Figure I. Two Demonstrations Highlighting the Extent to which People Overestimate their Perceptual Experience.

research question and future work will need to examine this issue with different paradigms and tasks (see Outstanding Questions).

Can these Mechanisms Explain the Richness of Experience?

Those who believe that visual awareness overflows mechanisms, such as attention and working memory, might say that our expanded notion of what observers can access still does not account for the richness of experience. One classic kind of evidence in support of the overflow argument comes from the partial report paradigms [97,98]. In these studies, participants encode items into working memory and are then quickly cued as to which particular items they should report. When cued in this way, performance is near ceiling. However, when no cue occurs, and participants must report the entire set, performance is worse. This finding has been cited as evidence of information overflowing cognitive access [6,10]. However, other researchers argue that the cue simply elevates previously unconscious information to consciousness [15,16]. Currently, there is no clear, uncontroversial way to empirically distinguish these interpretations, and so we focus the rest of our discussion on other types of information that may overflow access.

Even after considering ensemble statistics and scene perception, some people may still have the intuition that observers can see more than attention and working memory can capture. However, without clear empirical evidence to support this claim, it appears to be based purely on intuition. Should we trust this intuition? It is well established that observers systematically overestimate the richness of their own perception. People often believe that detecting changes in a change blindness experiment will be easy, and are surprised to find out that it is not [42,43]. Similarly, people do not realize how bad their acuity and color perception is in the periphery [1]. Box 2 presents two simple exercises that directly test the extent to which people are mistaken about their own perceptual experience.

What these exercises show is that it is easy to be wrong about the richness of perception, and scientists should be skeptical about claims that observers can see more than can be accessed. Instead, the claim that visual awareness overflows cognitive access must be supported by specific examples of visual input that can be consciously perceived without being attended, held in working memory, reported, or used to guide volitional action. Without specific evidence, there appears no good scientific reason to believe consciousness overflows cognition.

Concluding Remarks: Reconsidering the Focus of Consciousness Research

Many researchers have claimed that information cannot be consciously perceived without being accessed by higher-level cognitive functions, such as attention and working memory [13–18]. This view has been criticized for its inability to capture the richness of perceptual experience, given the strict capacity limits of these mechanisms evident in phenomena such as change blindness and inattentional blindness [7–11]. Critics of this view say that those who claim information must be accessed to be conscious believe that 'conscious perception is limited to the contents of visual working memory, roughly three or four things at a time in many standard paradigms' ([99] p. 445).

To the contrary, we argue that observers have access to considerably more information than just three or four items at a time. Instead, a handful of items are perceived with high fidelity, while the remainder of the world is represented as an ensemble statistic (or set of statistics). Those who link consciousness with higher-level cognitive function [13–18] need not believe that perception is sparse, with observers seeing only a few items at a time. Perception is undoubtedly rich, but this richness can be easily captured by cognitive mechanisms, such as attention and working memory. The focus of consciousness research should be on the nature of the visual information that is captured beyond the few high-fidelity objects that can be held in visual working memory, the nature and number of such summary statistics, and the capacity limits entailed in their extraction and representation.

Outstanding Questions

How are ensemble statistics represented in the brain? Are they formed across the same circuits involved in representing individual items? Or do noisy representations of individual items have to be read out by a higher-order node that forms a new, ensemble representation?

What are the attentional requirements of ensemble perception? Is it possible for ensemble statistics to be rendered inattentionally blind or go unnoticed due to the attentional blink? What type of attention is needed for ensemble perception? Can multiple statistics, summarizing different dimensions of the scene (i.e., average color, orientation, size, etc.), be established in parallel?

To what extent is ensemble perception used across sensory modalities besides vision? Do the same principles discovered in vision apply to other modalities?

Are higher-level social, physical, and action-based properties of scenes extracted quickly? Is different neural tissue engaged in extracting each of these kinds of information? To what extent does the subjective richness of perception result from the overlay of these higher-level physical, social, and action-based properties of a scene?

Supplementary Information

Supplementary information associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j. tics.2016.03.006.

References

1. Dennett, D.C. (1991) *Consciousness Explained,* Little Brown
2. Noë, A. (2002) Is the visual world a grand illusion? *J. Conscious. Stud.* 9, 1–12
3. Cohen, M.A. *et al.* (2012) The attentional requirements of consciousness. *Trends Cogn. Sci.* 16, 411–417
4. Luck, S.J. and Vogel, E.K. (2013) Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends Cogn. Sci.* 17, 391–400
5. Scimeca, J.M. and Franconeri, S.L. (2015) Selecting and tracking multiple objects. *Wiley Interdiscip. Rev. Cogn. Sci.* 6, 109–118
6. Block, N. (2011) Perceptual consciousness overflows cognitive access. *Trends Cogn. Sci.* 15, 567–575
7. Aru, J. and Bachman, T. (2013) Phenomenal awareness can emerge without attention. *Front. Behav. Neurosci.* 7, e891
8. Bronfmann, Z.Z. *et al.* (2014) We can see more than we can report: 'cost free' color phenomenality outside focal attention. *Psychol. Sci.* 25, 1394–1403
9. Koch, C. and Tsuchiya, N. (2007) Attention and consciousness: two distinct brain processes. *Trends Cogn. Sci.* 11, 16–22
10. Lamme, V. (2010) How neuroscience will change our view on consciousness. *Cogn. Neurosci.* 1, 204–220
11. Zeki, S. (2003) The disunity of consciousness. *Trends Cogn. Sci.* 7, 214–218
12. Dehaene, S. and Changeux, J.P. (2011) Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227
13. Baars, B. (1989) *A Cognitive Theory of Consciousness,* Cambridge University Press
14. Dehaene, S. (2014) *Consciousness and the Brain,* Viking Press
15. Cohen, M.A. and Dennett, D.C. (2011) Consciousness cannot be separated from function. *Trends Cogn. Sci.* 15, 358–364
16. Kouider, S. *et al.* (2010) How rich is consciousness? The partial awareness hypothesis. *Trends Cogn. Sci.* 14, 301–307
17. Lau, H. and Rosenthal, D. (2011) Empirical support for higher-order theories of conscious awareness. *Trends Cogn. Sci.* 15, 365–373
18. O'Regan, K. (2011) *Why Red Doesn't Sound Like a Bell,* Oxford University Press
19. Whitney, D. *et al.* (2014) From textures to crowds: multiple levels of summary statistical perception. In *The New Visual Neurosciences* (Wener, J.S. and Chalupa, L.M., eds), pp. 695–710, MIT Press
20. Alvarez, G.A. (2011) Representing multiple objects as an ensemble enhances visual cognition. *Trends Cogn. Sci.* 15, 122–131
21. Alvarez, G.A. and Cavanagh, P. (2004) The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychol. Sci.* 15, 106–111
22. Cowan, N. (2001) The magical number 4 in short-term memory: a reconsideration of memory storage capacity. *Behav. Brain Sci.* 24, 87–114
23. Drew, T. and Vogel, E.K. (2008) Neural measures of individual differences in selecting and tracking multiple moving objects. *J. Neurosci.* 28, 4183–4191
24. Luck, S.J. and Vogel, E.K. (1997) The capacity of visual working memory for features and conjunctions. *Nature* 390, 279–281
25. Franconeri, S.L. *et al.* (2007) How many locations can be selected at once? *J. Exp. Psychol. Hum. Percept. Perf.* 33, 1003–1012
26. Kanwisher, N. (2010) Functional specificity in the human brain: a window into the functional architecture of the mind. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1163–1170
27. Cant, J.S. and Xu, Y. (2012) Object ensemble processing in human anterior-medial ventral visual cortex. *J. Neurosci.* 32, 7685–7700
28. Cohen, M.A. *et al.* (2014) Processing multiple objects is limited by overlap in neural channels. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8955–8960
29. Cohen, M.A. *et al.* (2015) Visual awareness is limited by the representational architecture of the visual system. *J. Cogn. Neurosci.* 27, 2240–2252
30. Rensink, R.A. *et al.* (1997) To see or not to see: the need for attention to perceive changes in scenes. *Psychol. Sci.* 8, 368–373
31. O'Regan, J.K. *et al.* (1999) Change-blindness as a result of 'mudsplahes.'. *Nature* 398, 34–37
32. Simons, D.J. *et al.* (2000) Change blindness in the absence of a visual disruption. *Perception* 29, 1143–1154
33. Simons, D.J. and Chabris, C.F. (1999) Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception* 28, 1059–1074
34. Horrey, W.J. and Wickens, C.D. (2006) Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Hum. Factors* 48, 196–205
35. Kunar, M. *et al.* (2008) Telephone conversations impair sustained visual attention via a central bottleneck. *Psychon. Bull. Rev.* 15, 1135–1140
36. Suchow, J.W. *et al.* (2014) Terms of the debate on the format and structure of visual memory. *Attn. Percept. Psychophys.* 76, 2071–2079
37. Ma, W.J. *et al.* (2014) Changing concepts of working memory. *Nat. Neurosci.* 17, 347–356
38. Bays, P.M. and Husain, M. (2008) Dynamic shifts of limited working memory resources in human vision. *Science* 321, 851–854
39. Pylyshyn, Z.W. and Storm, R.W. (1988) Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spat. Vis.* 3, 179–197
40. Yantis, S. (1992) Multielement visual tracking: attention and perceptual organization. *Cogn. Psychol.* 24, 295–340
41. Howe, P.D.L. *et al.* (2008) Distinguishing between parallel and serial accounts of multiple object tracking. *J. Vis.* 10, 1–13
42. Levin, D.T. *et al.* (2000) Change blindness blindness: the metacognitive error of overestimating change-detection ability. *Vis. Cogn.* 7, 397–412
43. Scholl, B.J. *et al.* (2003) 'Change blindness' blindness: An implicit measure of a metacognitive error. In *Visual Metacognition: Thinking about Seeing* (Levin, D.T., ed.), pp. 145–164, MIT Press
44. Field, D.J. (1987) Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* 4, 2379–2394

45. Kersten, D. (1987) Predictability and redundancy of natural images. *J. Opt. Soc. Am. A* 4, 2395–2400

46. Geisler, W.S. (2008) Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol.* 59, 167–192

47. Rosenholtz, R. (2011) What your visual system seems where you are not looking. *Proc. SPIE* 7865, 786510

48. Alvarez, G.A. and Oliva, A. (2009) Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7345–7350

49. Brady, T.F. *et al.* (2011) A review of visual memory capacity: Beyond individual items and toward structured representations. *J. Vis.* 11, 1–34

50. Victor, J.D. and Conte, M.M. (2004) Visual working memory for image statistics. *Vis. Res.* 44, 541–546

51. Brady, T.F. and Tenenbaum, J.B. (2013) A probabilistic model of visual working memory: incorporating higher order regularities into working memory capacity estimates. *Psychol. Rev.* 120, 85–109

52. Cowan, N. (2001) The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behav. Brain Sci.* 24, 87–114

53. Brady, T.F. and Alvarez, G.A. (2011) Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. *Psychol. Sci.* 22, 384–392

54. Jiang, Y.V. *et al.* (2000) Organization of visual short-term memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 683–702

55. Vidal, J.R. *et al.* (2005) Relational information in visual short-term memory: the structural gist. *J. Vis.* 5, 244–256

56. Oliva, A. and Schyns, P. (1997) Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cogn. Psychol.* 34, 72–107

57. Schyns, P.G. and Oliva, A. (1994) From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychol. Sci.* 5, 195–200

58. Oliva, A. (2005) Gist of the scene. In *Neurobiology of Attention* (Itti, L. *et al.*, eds), pp. 251–257, Elsevier Press

59. Reninger, L.W. and Malik, J. (2004) When is scene identification just texture recognition? *Vis. Res.* 44, 2301–2311

60. Torralba, A. and Oliva, A. (2002) Depth estimation from image structure. *IEEE Pattern Anal. Mach. Intell.* 24, 1226–1238

61. Torralba, A. and Oliva, A. (2003) Statistics of natural images categories. *Network Comp. Neural* 14, 391–412

62. De Graef, P. (2005) Semantic effects on object selection in real-world scene perception. In *Cognitive Processes in Eye Guidance* (Underwood, G., ed.), pp. 213–235, Oxford University Press

63. Rayner, K. (1998) Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 85, 618–660

64. Henderson, J.M. (2003) Human gaze control in real-world scene perception. *Trends Cogn. Sci.* 7, 498–504

65. Potter, M.C. (1976) Short-term conceptual memory for pictures. *J. Exp. Psychol. Hum. Learn. Mem.* 2, 509–522

66. Fei-Fei, L. *et al.* (2007) What do we perceive in a glance of a real-world scene? *J. Vis.* 7, 1–29

67. Greene, M.R. *et al.* (2015) What you see is what you expect: rapid scene understanding benefits from prior experience. *Attn. Percept. Psychophys.* 77, 1239–1251

68. Greene, M.R. and Oliva, A. (2009) The briefest of glances: the time course of natural scene understanding. *Psychol. Sci.* 20, 464–472

69. Castelhano, M.S. and Henderson, J.M. (2008) The influence of color on scene gist. *J. Exp. Psychol. Hum. Percept. Perf.* 34, 660–675

70. Mack, A. and Rock, I. (1998) *Inattentional Blindness,* MIT Press

71. O'Regan, J.K. and Noë, A. (2001) A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 939–973

72. Oliva, A. *et al.* (2003) Top-down control of visual attention in object detection. *IEEE Int. Conf. Img. Proc.* 1, 253–256

73. Najemnik, J. and Geisler, W.S. (2005) Optimal eye movement strategies in visual search. *Nature* 434, 387–391

74. Neider, M.B. and Zelinski, G.J. (2006) Scene context guides eye movements during visual search. *Vis. Res.* 46, 614–621

75. Oliva, A. and Torralba, A. (2007) The role of context in object recognition. *Trends Cogn. Sci.* 11, 520–527

76. Wolfe, J. *et al.* (2011) Visual search in scenes involves selective and non-selective pathways. *Trends Cogn. Sci.* 15, 77–84

77. Rensink, R.A. (2000) The dynamic representation of scenes. *Vis. Cogn.* 7, 17–42

78. Kanwisher, N. (2010) Functional specificity in the human brain: a window into the functional architecture of the mind. *Proc. Natl. Acad. Sci. U.S.A.* 107, 11163–11170

79. Dilks, D. *et al.* (2011) Mirror-image sensitivity and invariance in object and scene processing pathways. *J. Neurosci.* 33, 11305–11312

80. Epstein, R. and Kanwisher, N. (1998) A cortical representation of the local visual environment. *Nature* 392, 598–601

81. Epstein, R.A. (2008) Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn. Sci.* 12, 388–396

82. Cant, J.S. and Goodale, M.A. (2007) Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cereb. Cortex* 17, 713–731

83. Peuskens, H. *et al.* (2004) Attention to 3-D shape, 3-D motion, and texture in 3-D structure from motion displays. *J. Cogn. Neurosci.* 16, 665–682

84. Steeves, J.K. *et al.* (2004) Behavioral and neuroimaging evidence for a contribution of color texture information to scene classification in a patient with visual agnosia. *J. Cogn. Neurosci.* 16, 955–965

85. Goodale, M.A. and Milner, A.D. (2004) *Sight Unseen,* Oxford University Press

86. Mendez, M.F. and Cherrier, M.M. (2003) Agnosia for scenes in topographagnosia. *Neuropsychologia* 41, 1387–1395

87. Cant, J.S. *et al.* (2015) Distinct cognitive mechanisms involved in the processing of single objects and object ensembles. *J. Vis.* 15, 1–21

88. Marois, R. *et al.* (2004) The neural fate of consciously perceived and missed events in the attentional blink. *Neuron* 41, 465–472

89. Stein, T. *et al.* (2009) The effect of fearful faces on the attentional blink is task independent. *Psychon. Bull. Rev.* 16, 104–109

90. Huang, L. (2015) Statistical properties demand as much attention as object features. *PLOS ONE* 10, e0131191

91. Feigenson, L. (2008) Parallel non-verbal enumeration is constrained by a set-based limit. *Cognition* 107, 1–18

92. Halberda, J. *et al.* (2006) Multiple spatially overlapping sets can be enumerated in parallel. *Psychol. Sci.* 17, 572–576

93. Poltoratski, S. and Xu, Y. (2013) The association of color memory and the enumeration of multiple spatially overlapping sets. *J. Vis.* 13, 1–11

94. Attarha, M. *et al.* (2014) Summary statistics of size: fixed processing capacity for multiple ensembles but unlimited processing capacity for single ensembles. *J. Exp. Psychol. Hum. Percept. Perf.* 40, 1440–1449

95. Attarha, M. *et al.* (2015) The capacity limitations of orientation summary statistics. *Attn. Percept. Psychophys.* 77, 1116–1131

96. Chong, S.C. and Treisman, A. (2005) Attentional spread in the statistical processing of visual displays. *Attn. Percept. Psychophys.* 67, 1–13

97. Sperling, G. (1960) The information available in brief visual presentation. *Psychol. Monogr.* 74, 1–29

98. Landman, R. *et al.* (2003) Large capacity storage of integrated objects before change blindness. *Vis. Res.* 43, 149–164

99. Block, N. (2014) Rich conscious perception outside focal attention. *Trends Cogn. Sci.* 18, 445–447

100. Freeman, J. and Simoncelli, E.P. (2011) Metamers of the ventral stream. *Nat. Neurosci.* 14, 1195–1201

101. Dakin, S.C. and Watt, R.J. (1997) The computation of orientation statistics from visual texture. *Vis. Res.* 37, 3181–3192

102. Baur, B. (2009) Does Stevens's power law for brightness extend to perceptual brightness averaging? *Psychol. Rec.* 59, 171–186

103. Watamaniuk, S.N.J. and Duchon, A. (1992) The human visual-system averages speed information. *Vis. Res.* 32, 931–941

104. Chong, S.C. and Treisman, A. (2003) Representation of statistical properties. *Vis. Res.* 43, 393–404

105. Haberman, J. and Whitney, D. (2007) Rapid extraction of mean emotion and gender from sets of faces. *Curr. Biol.* 17, R751–R753

106. de Fockert, J. and Wolfenstein, C. (2009) Rapid extraction of mean identity from sets of faces. *Q. J. Exp. Psychol.* 62, 1716–1722

107. Sweeny, T. and Whitney, D. (2014) Perceiving crowd attention: ensemble perception of a crowd's gaze. *Psychol. Sci.* 25, 1903–1913

108. Sweeny, T.D. *et al.* (2013) Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *J. Exp. Psychol. Hum. Percept. Perf.* 39, 329–337

109. Haberman, J. and Whitney, D. (2009) Seeing the mean: ensemble coding for sets of faces. *J. Exp. Psychol. Hum. Percept. Perf.* 35, 718–734

110. Albrecht, A.R. and Scholl, B.J. (2010) Perceptually averaging in a continuous visual world extracting statistical summary representations over time. *Psychol. Sci.* 21, 560–567

111. Haberman, J. *et al.* (2009) Averaging facial expression over time. *J. Vis.* 9, 1–13

112. Oliva, A. and Torralba, A. (2006) Building the gist of a scene: the role of global image features in recognition. *Progress Brain Res. Visual Percept.* 155, 23–36