

# 7

## SPATIAL STATISTICS IN PERCEPTION, LEARNING, AND NAVIGATION

*Kathryn N. Graves and Nicholas B. Turk-Browne*

### **Introduction**

The visual world around us, though rich and complex, is highly structured. We are often confronted with novelty, but just as often encounter some variation of things seen before. The very first time that you ever saw salt and pepper on a dinner table, you may not have immediately thought of them as paired together in a meaningful way. Now, however, whenever you see salt on a dining table you probably expect to see pepper nearby, and vice versa.

These underlying regularities in the things we experience provide an organizing principle by which we represent and remember our environment. The process of acquiring regularities operates across multiple dimensions of perception—from computing the summary statistics of a single exposure, for example rapidly estimating the average movement direction of a flock of birds, to extracting probabilities across a series of exposures, for example learning to expect where furniture and objects are located in a room. The flexible generation of these statistical representations allows us to acquire the rich structure of the world.

In this chapter, we will illustrate the pervasiveness of statistical processing in the mind and brain by exploring its influence on what we *perceive* in space in an instant, what we *learn* about space across time, and how we *navigate* space as we acquire our environment's underlying structure.

### **Spatial regularities in perception**

Imagine a crowd of people at a concert. If asked to calculate the average height of the crowd, how would you do it? Perhaps you'd measure and sum the height of each individual person one by one, a slow process. But what if you have less than a second to perform your calculation? In that case the only option is to make your best guess. As it turns out, your best guess is more accurate than you might think.

Humans are proficient at extracting underlying statistical information from visual input. This relies on *ensemble representations*—higher-order representations that collapse across individual items to describe features of a *set* of perceptual objects as a whole (Alvarez, 2011), such as their average or count. Here, we will focus primarily on perceptual averaging, but will return to the more general concept of ensemble representations when discussing limitations to this type of representation.

### ***Perceptual averaging***

When confronted with an array of lines of different lengths, people can compute the average length of those lines with striking accuracy (Miller & Sheldon, 1969). This computational feat of *perceptual averaging* is highly robust and operates over more than just the length dimension. Whether the average size of circles (Ariely, 2001), the center of mass of moving points (Alvarez & Oliva, 2008), the average orientation of Gabor patches (Parkes et al., 2001), or the average brightness of filled circles (Bauer, 2009), the human visual system rapidly acquires the underlying central tendency of various features and dimensions.

This averaging process requires only a brief exposure to the set. In one experiment, participants were asked to compute the average size of a set displayed for 1000, 100, or only 50 milliseconds. Across all three durations, participants displayed robust perceptual averaging, with minimal deterioration in accuracy even with only 50 milliseconds (Chong & Treisman, 2003). This finding underscores the automaticity of perceptual averaging.

Further, the visual system is not limited to summarizing basic physical features and textures. When people were presented with a series of faces ranging from “happy” to “sad” or “male” to “female”, they showed a bias for the average emotion or average gender of the set (Haberma & Whitney, 2007), even when the load was relatively high, with 16 faces in a set (Haberma & Whitney, 2009).

What is the role of the set constituents in perceptual averaging, and what happens to them in this process? A reasonable theory might be that the details of an individual target item are discarded during the calculation, as seems evident in the phenomenon of *crowding*. When multiple objects are presented simultaneously and in close proximity, visual discrimination of a single object is inhibited by the presence of this crowd (Levi, 2008). This suggests that individual objects are not represented when viewing an ensemble, yet they play a critical role in generating the statistical summary. Models of perceptual averaging argue that individual elements are not forgotten writ large but are combined and pooled in the human visual system (Parkes et al., 2001), out of which arises a summary statistic.

### ***Robustness to noise***

Regularities refer to the aspects of a display that are shared, but these are usually embedded among other idiosyncratic or noisy aspects. Data points that represent extremes within a set, for example, may be less informative of the underlying statistical structure. When confronted with this minimally informative data, studies show

that human learners display robust perceptual averaging. Just as a statistical analysis might exclude outliers, people discard noisy data from their computation instead of simply integrating all possible data points (de Gardelle & Summerfield, 2011).

Importantly, outliers are not the only source of noise. The items in a set may vary on multiple dimensions, not all of which are relevant for a given computation. If you are estimating the average height of individuals in a crowd, for example, the proximity of individuals should not influence your calculation. In one experiment, researchers sought to determine how such irrelevant dimensions influence perceptual averaging, predicting that a sufficiently robust perceptual averaging process would separate out the various dimensions and operate selectively over the relevant one. The participants were instructed to compare the mean sizes of two arrays presented side by side. However, either or both the density and number of the objects in each array could be varied independent of the mean. Estimates of mean size were unaffected when one of these dimensions varied (Chong & Treisman, 2005), demonstrating robustness. However, performance was impaired when both varied concurrently, suggesting that these dimensions are processed even when ignored in isolation, and that there is a limit on robustness when faced with extreme variability. This could reflect an influence of these dimensions on the mean computation or response-level conflict in which density or number attributes are inadvertently substituted for the mean and favor selection of the incorrect array.

### ***Distributional complexity***

Beyond noise and variability, the shape of the underlying distribution can vary across features and time. The most straightforward structure from which to extract a central tendency is a Gaussian distribution. However, in the real world, sets do not always follow that shape.

Returning to our height estimation analogy, imagine you're confronted with two groups of people instead of one—a group of adults and a group of children. Even if the average height in the adult group followed a normal distribution, the full set is bimodal, with peaks for both the average adult height and average child height. Incredibly, perceptual averaging can handle this complex statistical structure. One experiment asked people to determine which average size was larger between a pair of two side-by-side distributions. The shape of each distribution in the pair was either Gaussian, uniform, bimodal, or homogenous, and all possible combinations of shape pairs were tested. Across each shape combination, people responded to the two distributions with accuracy comparable to that of a single distribution (Chong & Treisman, 2003).

In sum, perceptual averaging is pervasive, enduring despite noisy environments, non-normal set distributions, and the presence of multiple sets at once.

### ***Reduced attention***

Rarely in the real world do we focus one hundred percent of our attention on a single task. Even as you're reading this book, you may also be unconsciously

monitoring your cell phone notifications and trying to adjust to a more comfortable sitting position. In the real world, we often attend to multiple things at once: The average height of the people in a crowd, the slowly advancing rain clouds in the distance, which pocket our wallet is in, and so on. Under reduced attention conditions such as these, individuals are worse at remembering spatial locations of individual items in a display; however, they are highly accurate at reporting the center of mass of the set (Alvarez & Oliva, 2009).

Further, in some cases attending to multiple things at once is optimal for perceptual averaging. In one study, researchers argued that the averaging process is more efficient under conditions in which attention is distributed over a display, instead of focused. They found that perceptual averaging during a simultaneous distributed-attention task was comparable to that of a pure perceptual averaging task with no additional demand (Chong & Treisman, 2005).

Thus, perceptual averaging is well-adapted for use in the real world. When an individual is unable to devote their full attention to the set in front of them, or is trying to do multiple things at once, they will still likely acquire the underlying structure of the set.

### ***Development***

We have provided ample evidence that perceptual averaging is a rapid, automatic, and durable process. However, as a form of statistical calculation, it may develop like other analytical skills that improve significantly during childhood (Kolkman et al., 2013). Indeed, children show noticeable deficits and variations in strategy when compared to adults in performing perceptual averaging. In one study, four- and five-year-old children were proficient in estimating which set in a pair of arrays contained a larger mean size, but were still less accurate than adults (Sweeny et al., 2015).

What drives these developmental differences? A study of average motion perception across development used computational modeling to determine how statistical computations differ between children and adults. Child behavior was better fit by models that averaged fewer of the perceptually available motion directions, suggesting that development may relate to the nature of sampling, with larger samples drawn as a function of age (Manning et al., 2014). More generally, then, an age-related increase in the number of data points sampled from a set may contribute to greater accuracy of perceptual averaging.

However, differences across development may also be a function of qualitatively different strategies used by children and adults. When participants from both age groups were asked to locate the middle of displays of randomly generated dot-clouds, at a low set size children's performance reflected a strategy of "joining up the dots" and estimating the center of gravity of the subsequent shape. This was considered a qualitative difference when compared to adults' strategy of arithmetically averaging the individual points (Jones & Dekker, 2018). Thus, a combination of computational refinement and qualitative shift in strategy may preempt the robust perceptual averaging capacities displayed in human adults.

### ***Limitations to ensemble representation***

Although robust to noise, variability, and inattention, reported *confidence* in statistical computations is nevertheless subject to the amount of environmental noise over which those computations are conducted (Zylberberg et al., 2012).

More generally, system constraints impose limits on the quantity of data over which ensemble representations operate. *Visual working memory* (VWM), a temporary storage of accessibly formatted visual information (Baddeley, 1992), underlies the generation of ensemble representations. This storage system has limited capacity, resulting in downstream limits on the number of representations that can be generated at one time. For example, when asked to enumerate multiple sets of items simultaneously, people could do so for up to three sets, but no more (Feigenson, 2008; Halberda et al., 2006). Thus, VWM imposes a hard cap on the number of statistical structures that can be acquired at one time.

Additionally, while ensemble representations in visual memory enable highly accurate reporting of the central tendency of perceived displays, this can come at the cost of memory for individual items in the set. Even when some items are stored in memory, these memories are biased towards the central tendency (Brady & Alvarez, 2011). Thus, although ensemble representations provide a helpful heuristic for quickly summarizing the perceptually rich visual world, generation of these representations is not without consequences.

Nonetheless, ensemble representations are a robust mechanism by which people represent and remember features of the implicit structure of their immediate visual environment. Even though peak proficiency with perceptual averaging is not reached until adulthood, its computational power and flexibility (but also its limitations) are evident in the examples we have reviewed thus far. However, these scenarios have involved displays of static statistical distributions. Critically, the real world is anything but static. Rather, in many instances structure becomes apparent over time. Thus, an useful statistical mechanism must be able to acquire underlying statistics not just from the current visual experience, but by aggregating across a series of experiences.

### ***Spatial regularities in learning***

Say that you went to a big-name grocery store in your city and noticed that the oranges were displayed on the far-right area of the produce section. On a different day, you went to another grocery store location and noticed that the oranges were again to the far right. You have the same experience at another location, and another. Later you enter another grocery store at yet another different location. Where would you first look for oranges?

At this point, you would likely say the far-right area of the produce section. However, you did not start out having this prediction by default—the regularity on which you are mounting your prediction emerged gradually, over shopping trips. As mentioned before, ensemble statistical processing is highly flexible. As with acquiring structure from a single perceptual array, the visual system can acquire spatial statistics across time by linking sequential visual experiences.

### ***Distributions in time***

We previously discussed perceptual averaging in the context of static displays. However, spatial statistics can also emerge over time. In one study, participants viewed displays of dots whose motion was drawn from a distribution, where a subset of directions occurred with high frequency across trials (Chalk et al., 2010). When asked to detect the motion direction of a dot display, reports were significantly biased towards the most frequently presented directions. Further, on trials where no motion signal was presented (pure noise), participants “hallucinated” stimuli moving in the most frequent direction. Thus, these learned regularities created spatial statistical representations so ingrained that they altered perceptual experience.

### ***Probability cueing***

Extraction of structure across time extends beyond the central tendency. As with our grocery store example, a visual stimulus might be more likely to occur in one location than another, based on prior experience. That underlying probability structure serves as a spatial cue to what will appear in the future. In one study of this *probability cueing*, participants searched for target letters within a letter display, where different target letters appeared with different frequencies at specific locations (Hoffmann & Kunde, 1999). Participants were significantly faster at detecting target letters when they appeared in their high-frequency locations, suggesting that over the course of learning, they established expectations based on the underlying probabilities. This computation is flexible, operating over absolute (Shaw & Shaw, 1977) and relative spatial positions (Miller, 1988). It is also highly persistent: People who learned target location probabilities during one study session showed a preserved spatial bias a week later, even after the probability structure changed such that the target’s likely location was now evenly distributed across all possible locations (Jiang et al., 2013).

Probability cueing functions not only to highlight where to attend, in the context of target locations, but also where not to attend, in the context of suppressing locations of distractors. Just as people are faster to respond to target items that appear in high-probability locations, they are better at ignoring distracting, non-target items via the same statistical property (Goschy et al., 2014). Though like cueing of target locations, previous work has shown that distractor suppression is a separate calculation (Failing et al., 2019). Additionally, this suppression effect is uniquely a function of probability cueing. One study tried and failed to elicit similar suppression by explicitly cueing each distractor location, instead of probability functioning as an implicit cue (Wang & Theeuwes, 2018). This process also scales with the salience of the distractor, such that the degree of attention capture by a given distractor predicts the extent of subsequent suppression (Failing & Theeuwes, 2020).

These phenomena demonstrate the power of probability cueing. This process imparts no physical change on the perceived stimuli, but instead tangibly alters the salience of the spatial locations where those stimuli appear. However, directly cue-

ing a target or distractor is not the only way to generate expectations that facilitate performance. The context in which these items occur also holds predictive power.

### ***Contextual cueing***

Once you stepped foot inside the last grocery store, probability cueing would have led you to look to the far right for your oranges, as previous experience determined that to be a high-probability location for oranges. However, this is not the only form of statistics-based cueing that led you to look for oranges in this location. Other cues as you were walking down the street (e.g., the shape and location of the building) and entered the store (e.g., the arrangement of aisles and open areas) also predicted where you were going to find oranges.

This *contextual cueing* occurs when the locations, configuration, and identities of some stimuli inform where and what another stimulus will be. In a classic study, participants were asked to quickly respond to the orientation of a rotated target “T” that appeared amidst a configuration of distractor “L”s (Chun & Jiang, 1998). Unbeknownst to the participants, half of the L configurations were repeated across trials, and the rest were novel and did not repeat. In the repeated configurations, the target T appeared in a specific location relative to the Ls. A response time benefit emerged for the repeated versus novel configurations over the course of the task, indicating that participants associated repeated configurations with high-probability locations for a target to appear. Of note, learning in contextual cueing occurs at multiple levels, as people acquire both configural associations and more granular associations between targets and local or individual distractors (Jiang & Wagner, 2004; Brady & Chun, 2007).

Probability thus far has been identified as a mechanism by which task-relevant spatial locations are highlighted—whether for expecting a target or suppressing a distractor, and whether as a function of the stimulus location or the contextual configuration. However, the power of probability to influence expectations is pervasive, and functions not only in these domains but also for stimulus-stimulus relationships.

### ***Visual statistical learning***

Beyond cueing the location or identity of individual stimuli, probabilities can link two or more stimuli across space and time. If, every time you saw oranges to the far right in the produce section, you saw limes next to them, finding limes might lead to an expectation to see oranges. These *co-occurrences* between visual inputs across experiences provide another means for learning the structure of the environment. Here the role of probability is to generate associations between stimuli based on their regularities, what is known as *visual statistical learning*.

An initial investigation of visual statistical learning had participants passively view a series of unique and complex visual scenes (Fiser & Aslin, 2001). Unbeknownst to them, these scenes were composed of pairs of objects that had a consistent spatial relation. Even though those relationships were never segmented

or otherwise explicitly indicated during learning, participants were later able to discriminate true from foil pairs at test. Participants acquire spatial statistical relationships at multiple scales, from the frequency and absolute position of individual shapes, to position-independent pair configurations and conditional probabilities of co-occurrence, to higher-order structures (Orbán et al., 2008). Thus, visual statistical learning may involve more than one type of computation, which collectively represent regularities experienced over time.

Acquiring spatial regularities comes at the cost of the specific memory for individual items within a configuration. When asked to recall the individual locations of spatially paired items, people remember them as being physically closer together in space than they were (Yu & Zhao, 2018). This parallels findings in perceptual averaging, where memories for the sizes of individual items in a display were biased towards the mean size of the set (Brady & Alvarez, 2011).

The parallel impairment of item memory by both visual statistical learning and perceptual averaging suggests that these processes are related. Indeed, one study found evidence of competition between perceptual averaging and visual statistical learning, such that acquisition of one statistic impeded acquisition of the other (Zhao et al., 2011). This mutual interference suggests that, despite the slower multi-experience timescale of visual statistical learning as compared to single-experience perceptual averaging, these processes may depend on overlapping statistical computations.

### ***Development***

Thus far we have discussed the development of statistical perception, emphasizing refinements to this process over time. Divergent from the developmental trajectory of perceptual averaging, probability cueing does not appear to require a protracted maturation period. Children as young as six years old demonstrated the cueing effect, at a level comparable to their adult counterparts (Yang & Song, 2020). Likewise, children and adults perform equivalently in contextual cueing tasks, at least when the ratio of repeated versus new displays is high (Yang & Merrill, 2015).

The early development of statistical abilities is especially clear for visual statistical learning, a precocious ability evident in newborns (Bulf et al., 2011) and infants as young as two months old (Kirkham et al., 2002). This immediate availability illustrates the foundational importance of visual statistical learning in providing an initial basis for our knowledge about the structure and contents of our environment. Nevertheless, statistical learning abilities continue to develop through childhood (Shufaniya & Arnon, 2018). How this developmental trajectory aligns with that of perceptual averaging remains an open question.

### ***Neural underpinnings***

How might this sophisticated and foundational learning be instantiated in the brain? In determining the neural basis of spatial statistical extraction, functional magnetic resonance imaging (fMRI), magnetoencephalogram (MEG), and lesion



studies have all implicated a common region, the hippocampus. An early lesion study found that amnesic patients with damage to the medial temporal lobe were impaired on a contextual cueing task (Chun & Phelps, 1999). Critically, their impairments were specific to implicit learning of the underlying statistical structure. Both amnesics and controls improved generally in learning the task over time, but only controls showed additional facilitation with the repeated compared to novel displays. Later fMRI work found reliably lower hippocampal activity in response to repeated versus novel context displays, showing that this region differentiated stimuli based on acquired structure (Greene et al., 2007). The hippocampus exerts influence over the visual cortex via this extracted structure, directing attention to high-likelihood target locations (Stokes et al., 2012). MEG has emphasized the role of the hippocampus particularly early in learning, with one study showing peak hippocampal activation during the first few trials of a spatial contextual cueing task (Spaak & de Lange, 2020).

In visual statistical learning, hippocampal activity reflects acquired statistical relationships in learning-induced changes to the neural representations of individual stimuli. When initially unpaired items become temporally associated via statistical learning, the patterns of hippocampal activity they evoke individually become more similar (Schapiro et al., 2012). The hippocampus similarly supports statistical learning in space: When acquiring spatial relationships between pairs of stimuli, hippocampal activity predicts the extent of behavioral learning (Karuz et al., 2017).

Statistical learning is thought to rely on the relational binding mechanism of the hippocampus, by which associations are formed rapidly and without conscious awareness (Ryan et al., 2000). In one study, participants saw a series of faces overlaid on scenes. At test, they were then shown the studied scenes, overlaid following a brief delay with three faces. One was the paired face from practice, with the others having been paired with other scenes. Participants responded by indicating the paired face. Hippocampal activity during the delay period tracked the association, as revealed by eye movements to paired vs. unpaired faces, even when participants failed to identify the correct face consciously (Hannula & Ranganath, 2009).

The implicitness of relational binding aligns with statistical learning, supporting the automaticity of this process and downplaying the role of consciousness in hippocampal processing, a traditional hallmark of its key role in episodic memory (Henke, 2010). Via relational binding, the hippocampus flexibly extracts structure across modalities and represents statistics.

### **Spatial statistics in navigation**

We have discussed the cognitive and neural mechanisms that allow us to perceive and learn spatial structure. However, most of the studies are constrained with respect to how we encounter spatial statistics in the real world: Tasks take place in a laboratory setting, from a single egocentric viewpoint, on a 2-D display, and by stationary participants. This is not representative of naturalistic visual experience in which viewpoints change and humans actively move through their environments.

Thus, a major question remains as to how these processes translate from sitting in a lab to navigating the real world. Although this mystery has not yet been solved, there are clues that link statistical learning to navigation in the presence of a shared neural mechanism, and evidence of spatial distributional learning in other contexts.

### ***Hippocampal mechanisms of navigation***

Spatial navigation, the foundational and essential behavior that allows organisms to move through and change their environment, is supported by the hippocampus in many species, including cats (Coleman & Lindsley, 1975), rats (Davidson et al., 2009), and primates (Gulli et al., 2020). When navigating a novel environment, specialized “place cells” fire in distinct locations such that the firing pattern creates an allocentric “map” of the space (O’Keefe & Dostrovsky, 1971). At the level of neuronal ensemble dynamics, hippocampal theta oscillations are stronger during navigation than during rest (Miller et al., 2018) and may encode environmental novelty (Pu et al., 2017). When it comes to remembering navigated paths, this theta rhythm in concert with higher-frequency sharp wave ripples support the “replay” of experienced trajectories, stabilizing the learned spatial map (Davidson et al., 2009). Thus, it is evident that the hippocampus performs multiple essential functions during navigation. How can we reconcile this role of the hippocampus in navigation with its role described above in statistical learning?

### ***Statistical spatial representations***

Given the ubiquity of statistical learning, we now explore how this mechanism may extend to and support navigation-related processes. Although learning a pair structure or computing an average line orientation are not overtly critical during navigation, there is evidence that spatial statistics ground representations used during navigation. Early models of navigation in the rodent hippocampus attributed place cell firing during learning of a novel environment to the construction of a navigational map (Blum & Abbott, 1996). It is not the discrete firing of a single place cell in a single location, but the coordinated firing of cells throughout a learned environment, that support the aggregated map representation. Further, this representation is not immediately accessible upon first exposure to a novel environment, just as spatial or temporal regularities are not learned from a single exposure to one item in a series. Instead, this map representation comes online as a function of linking individual locations together during exploration of a new environment.

Additionally, relational binding, a building block of statistical learning, extends beyond binding objects to scene images and may be essential in learning the locations of landmarks relevant to navigation. An analogous process of *object-place* association formation binds objects to locations in space. One study showed the criticality of the hippocampus in this process by lesioning rats after they had acquired an association. These rodents showed significant initial deficits in retaining the association when compared to rats with cortical lesions, indicating that the bound representation was stored in the hippocampus (Gilbert & Kesner, 2004).

Given that landmark use is predicated on representing objects in space (Deshmukh & Knierim, 2013), this associative mechanism that preempts statistical learning seems additionally essential for building representations of features of navigable environments.

### ***Learning in navigable space***

Even though spatial representations are essential for effective navigation, it remains unclear what role statistical learning, as explored above, plays in establishing and linking these representations. Work with rodents provides some initial insight that statistical learning permeates this representational space. For example, when rodents learn a series of rooms with partially overlapping spatial layouts, a subset of hippocampal neurons will consistently fire across all rooms (Leutgeb et al., 2004). In other words, the hippocampus has a specialized response to environmental regularities during navigation.

On a longer timescale of days to weeks, the generation of *schema*, or prior knowledge, of spatial relationships indicates another instantiation of statistical learning during navigation. After rodents were trained to associate a series of foods with different spatial locations over the course of six weeks, they more rapidly learned a new food in a new location that fit within the learned framework (Tse et al., 2011). Similarly, across a multi-day Morris Water Maze experiment resembling a protracted perceptual averaging task, rodents learned to find one hidden location per day from a set of normally distributed spatial locations. When their learning was tested 30 days later, their search behavior more closely resembled the underlying distribution from which the locations were drawn than the true locations themselves, suggesting that, over the course of consolidation, the rodents had extracted this pattern (Richards et al., 2014).

Although these paradigms implicate neocortical regions as the locus of spatial pattern representation (e.g., medial prefrontal cortex), pattern acquisition on a shorter timescale may be governed by the hippocampus (Schapiro et al., 2017). Similar spatial pattern extraction has been established in humans using a water maze paradigm, with one key difference: Instead of spreading learning across multiple days, people learned a distribution of locations within one session—a timescale more akin to visual statistical learning tasks (Graves et al., 2020). Within a single session, as participants acquired individual locations from a normally distributed set, their search behavior for new locations became biased toward the set mean. This behavior and timescale, though too gradual for perceptual averaging, are reminiscent of the biases found during spatial statistical learning. This stands in opposition to the much slower timescale of schema generation, and serves as a bridge between statistical learning in the lab and in the real world. It may be that the same mechanisms that support statistical learning in 2-D operate over navigable space. For example, the mechanism described earlier whereby statistical learning leads two associated objects to become represented more similarly in the hippocampus (Schapiro et al., 2012) has been shown when two locations have been linked through navigation (Deuker et al., 2016).

## Conclusion

In this chapter, we outlined the pervasiveness with which spatial statistics influence our representation of the world around us. From perception of ensemble statistics, to learning of regularities over time, to acquisition of structure in the real world, statistics alter our perception of space, our knowledge about the environment, and our subsequent movements and behavior. As such, this chapter serves as a point of convergence between three seemingly distinct literatures—perception, learning, and navigation—by recognizing their commonalities in terms of shared cognitive and neural mechanisms.

There is emerging evidence that spatial statistical learning enables us to acquire the regularities in our navigable environment. Future work is needed to confirm the role of the hippocampus in extracting statistics from the real world, as has been shown in the lab. Given the multimodal nature of statistical learning, investigations should translate spatial statistical learning phenomena from two to three dimensions, so we can better understand what statistics and patterns can be acquired readily as we navigate the world. New techniques may be needed as well to obtain neural measures during navigation, as the brain responds differently to virtual and real-world movement (Bohbot et al., 2017), including the development of wearable devices and brain implants (Stangl et al., 2021).

## 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 Sciences*, 106(18), 7345–7350.
- Ariely, D. (2001). Seeing Sets: Representation by Statistical Properties. *Psychological Science*, 12(2), 157–162.
- Baddeley, A. (1992). Working Memory. *Science*, 255(5044), 556–559.
- Bauer, B. (2009). Does Stevens's Power Law for Brightness Extend to Perceptual Brightness Averaging? *The Psychological Record*, 59, 171–185.
- Blum, K. I., & Abbott, L. F. (1996). A Model of Spatial Map Formation in the Hippocampus of the Rat. *Neural Computation*, 8(1), 85–93.
- Bohbot, V. D., Copara, M. S., Gotman, J., & Ekstrom, A. D. (2017). Low-frequency theta oscillations in the human hippocampus during real-world and virtual navigation. *Nature Communications*, 8, 14415.
- Brady, T. F., & Alvarez, G. A. (2011). Hierarchical Encoding in Visual Working Memory: Ensemble Statistics Bias Memory for Individual Items. *Psychological Science*, 22(3), 384–392.
- Brady, T. F., & Chun, M. M. (2007). Spatial constraints on learning in visual search: Modeling contextual cuing. *Journal of Experimental Psychology: Human Perception and Performance*, 33(4), 798–815.
- Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, 121(1), 127–132.
- Chalk, M., Seitz, A. R., & Seriès, P. (2010). Rapidly learned stimulus expectations alter perception of motion. *Journal of Vision*, 10(8), 2.

- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, 43(4), 393–404.
- Chong, S. C., & Treisman, A. (2005). Statistical processing: computing the average size in perceptual groups. *Vision Research*, 45(7), 891–900.
- Chun, M. M., & Jiang, Y. (1998). Contextual Cueing: Implicit Learning and Memory of Visual Context Guides Spatial Attention. *Cognitive Psychology*, 36(1), 28–71.
- Chun, M. M., & Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neuroscience*, 2, 844–847.
- Coleman, J. R., & Lindsley, D. B. (1975). Hippocampal electrical correlates of free behavior and behavior induced by stimulation of two hypothalamic-hippocampal systems in the cat. *Experimental Neurology*, 49(2), 506–528.
- Davidson, T. J., Kloosterman, F., & Wilson, M. A. (2009). Hippocampal Replay of Extended Experience. *Neuron*, 63(4), 497–507.
- De Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences*, 108(32), 13341–13346.
- Deshmukh, S. S., & Knierim, J. J. (2013). Influence of local objects on hippocampal representations: Landmark vectors and memory. *Hippocampus*, 23(4), 253–267.
- Deuker, L., Bellmund, J. L., Schröder, T. N., & Doeller, C. F. (2016). An event map of memory space in the hippocampus. *Elife*, 5, e16534.
- Failing, M., & Theeuwes, J. (2020). More capture, more suppression: Distractor suppression due to statistical regularities is determined by the magnitude of attentional capture. *Psychonomic Bulletin & Review*, 27, 86–95.
- Failing, M., Wang, B., & Theeuwes, J. (2019). Spatial suppression due to statistical regularities is driven by distractor suppression not by target activation. *Attention, Perception, & Psychophysics*, 81, 1405–1414.
- Feigenson, L. (2008). Parallel non-verbal enumeration is constrained by a set-based limit. *Cognition*, 107(1), 1–18.
- Fiser, J., & Aslin, R. N. (2001). Unsupervised Statistical Learning of Higher-Order Spatial Structures from Visual Scenes. *Psychological Science*, 12(6), 499–504.
- Gilbert, P. E., & Kesner, R. P. (2004). Memory for objects and their locations: The role of the hippocampus in retention of object-place associations. *Neurobiology of Learning and Memory*, 81(1), 39–45.
- Goschy, H., Bakos, S., Müller, H. J., & Zehetleitner, M. (2014). Probability cueing of distractor locations: Both intertrial facilitation and statistical learning mediate interference reduction. *Frontiers in Psychology*, 5, 1195.
- Graves, K. N., Antony, J. W., & Turk-Browne, N. B. (2020). Finding the pattern: On-Line Extraction of Spatial Structure During Virtual Navigation. *Psychological Science*, 31(9), 1183–1190.
- Greene, A. J., Gross, W. L., Elsinger, C. L., & Rao, S. M. (2007). Hippocampal differentiation without recognition: An fMRI analysis of the contextual cueing task. *Learning & Memory*, 14, 548–553.
- Gulli, R. A., Duong, L. R., Corrigan, B. W., Doucet, G., Williams, S., Fusi, S., & Martinez-Trujillo, J. C. (2020). Context-dependent representations of objects and space in the primate hippocampus during virtual navigation. *Nature Neuroscience*, 23, 103–112.
- 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.
- Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple Spatially Overlapping Sets Can Be Enumerated in Parallel. *Psychological Science*, 17(7), 572–576.
- Hannula, D. E., & Ranganath, C. (2009). The Eyes Have It: Hippocampal Activity Predicts Expression of Memory in Eye Movements. *Neuron*, 63(5), 592–599.
- Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews Neuroscience*, 11, 523–532.

- Hoffmann, J., & Kunde, W. (1999). Location-specific target expectancies in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 25(4), 1127–1141.
- Jiang, Y.V., Swallow, K. M., Rosenbaum, G. M., & Herzig, C. (2013). Rapid acquisition but slow extinction of an attentional bias in space. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), 87–99.
- Jiang, Y., & Wagner, L. C. (2004). What is learned in spatial contextual cuing—configuration or individual locations? *Perception & Psychophysics*, 66, 454–463.
- Jones, P. R., & Dekker, T. M. (2018). The development of perceptual averaging: learning what to do, not just how to do it. *Developmental Science*, 21(3), e12584.
- Karuz, E. A., Emberson, L. L., Roser, M. E., Cole, D., Aslin, R. N., & Fiser, J. (2017). Neural Signatures of Spatial Statistical Learning: Characterizing the Extraction of Structure from Complex Visual Scenes. *Journal of Cognitive Neuroscience*, 29(12), 1963–1976.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, 83(2), B35–B42.
- Leutgeb, S., Leutgeb, J. K., Treves, A., Moser, M. B., & Moser, E. I. (2004). Distinct Ensemble Codes in Hippocampal Areas CA3 and CA1. *Science*, 305(5688), 1295–1298.
- Kolkman, M. E., Kroesbergen, E. H., & Leseman, P. P. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learning and Instruction*, 25, 95–103.
- Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research*, 48(5), 635–654.
- Manning, C., Dakin, S. C., Tibber, M. S., & Pellicano, E. (2014). Averaging, not internal noise, limits the development of coherent motion processing. *Developmental Cognitive Neuroscience*, 10, 44–56.
- Miller, J. (1988). Components of the location probability effect in visual search tasks. *Journal of Experimental Psychology: Human Perception and Performance*, 14(3), 453–471.
- Miller, A. L., & Sheldon, R. (1969). Magnitude estimation of average length and average inclination. *Journal of Experimental Psychology*, 81(1), 16–21.
- Miller, J., Watrous, A. J., Tsitsiklis, M., Lee, S. A., Sheth, S. A., Schevon, C. A., ... & Jacobs, J. (2018). Lateralized hippocampal oscillations underlie distinct aspects of human spatial memory and navigation. *Nature Communications*, 9, 2423.
- O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34, 171–175.
- Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences*, 105(7), 2745–2750.
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4, 739–744.
- Pu, Y., Cornwell, B. R., Cheyne, D., & Johnson, B. W. (2017). The functional role of human right hippocampal/parahippocampal theta rhythm in environmental encoding during virtual spatial navigation. *Human Brain Mapping*, 38(3), 1347–1361.
- Richards, B. A., Xia, F., Santoro, A., Husse, J., Woodin, M. A., Josselyn, S. A., & Frankland, P. W. (2014). Patterns across multiple memories are identified over time. *Nature Neuroscience*, 17, 981–986.
- Ryan, J. D., Althoff, R. R., Whitlow, S., & Cohen, N. J. (2000). Amnesia is a Deficit in Relational Memory. *Psychological Science*, 11(6), 454–461.
- Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of Object Representations in the Human Medial Temporal Lobe Based on Temporal Regularities. *Current Biology*, 22(17), 1622–1627.
- Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: a neural network modeling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372: 20160049.
- Shaw, M. L., & Shaw, P. (1977). Optimal allocation of cognitive resources to spatial locations. *Journal of Experimental Psychology: Human Perception & Performance*, 3(2), 201–211.

- Shufaniya, A., & Arnon, I. (2018). Statistical Learning Is Not Age-Invariant During Childhood: Performance Improves With Age Across Modality. *Cognitive Science*, 42(8), 3100–3115.
- Spaak, E., & de Lange, F. P. (2020). Hippocampal and Prefrontal Theta-Band Mechanisms Underpin Implicit Spatial Context Learning. *Journal of Neuroscience*, 40(1), 191–202.
- Stangl, M., Topalovic, U., Inman, C. S., Hiller, S., Villaroman, D., Aghajan, Z. M., Christov-Moore, L., Hasulak, N. R., Rao, V. R., Halpern, C. H., Eliashiv, D., Fried, I., & Suthana, N. (2021). Boundary-anchored neural mechanisms of location-encoding for self and others. *Nature*, 589(7842), 420–425. <https://doi.org/10.1038/s41586-020-03073-y>
- Stokes, M. G., Atherton, K., Patai, E. Z., & Nobre, A. C. (2012). Long-term memory prepares neural activity for perception. *Proceedings of the National Academy of Sciences*, 109(6), E360–E367.
- Sweeny, T. D., Wurnitsch, N., Gopnik, A., & Whitney, D. (2015). Ensemble perception of size in 4–5-year-old children. *Developmental Science*, 18(4), 556–568.
- Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., ... & Morris, R. G. (2011). Schema-Dependent Gene Activation and Memory Encoding in Neocortex. *Science*, 333(6044), 891–895.
- Wang, B., & Theeuwes, J. (2018). How to inhibit a distractor location? Statistical learning versus active, top-down suppression. *Attention, Perception, & Psychophysics*, 80, 860–870.
- Yang, Y., & Merrill, E. C. (2015). The impact of signal-to-noise ratio on contextual cueing in children and adults. *Journal of Experimental Child Psychology*, 132, 65–83.
- Yang, Y., & Song, Q. (2021). Visual statistical learning in children and adults: evidence from probability cueing. *Psychological Research*, 85, 2911–2921.
- Yu, R. Q., & Zhao, J. (2018). Object representations are biased toward each other through statistical learning. *Visual Cognition*, 26(4), 253–267.
- Zhao, J., Ngo, N., McKendrick, R., & Turk-Browne, N. B. (2011). Mutual interference between statistical summary perception and statistical learning. *Psychological Science*, 22(9), 1212–1219.
- Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in Integrative Neuroscience*, 6, 79.