401
Chapter V. (Part II) Adventures into *terra incognita*
403
404   Ascending through the hierarchy of cortical computations, leaving
405   primary visual cortex, we reach the fascinating and bewildering cortical areas that
406   bridge low-level visual features into the building blocks of perception. In primary
407   visual cortex there are neurons that respond selectively to lines of different
408   orientation (**Chapter V, Part I**) (Hubel & Wiesel 1959, Hubel & Wiesel 1968). At
409   the other end of the visual hierarchy, there are neurons that respond selectively
410   to complex shapes such as faces, as we will discuss in **Chapter VI**. In between,
411   there is a large expanse of cortex involved in the magic transformations that take
412   oriented lines into complex shapes. How do we go from oriented lines to
413   recognizing faces and cars and other fancy shapes (**Figure 5.5**)? Despite heroic
414   efforts by a talented cadre of investigators to scrutinize the responses between
415   primary visual cortex and the highest echelons of inferior temporal cortex, this
416   part of cortex remains *terra incognita* in many ways. Visual information flows
417   along the ventral visual stream from primary visual cortex into areas V2, V4,
418   posterior and anterior parts of inferior temporal cortex. The cortical real estate
419   between V2 and inferior temporal cortex composes a mysterious, seductive,
420   controversial and fascinating ensemble of neurons whose functions remain
421   unclear and are only beginning to be deciphered.

Figure 5.5: Through the cascade of computations along the ventral visual stream, the brain can convert preferences for simple stimulus properties such as orientation tuning into complex features such as faces.
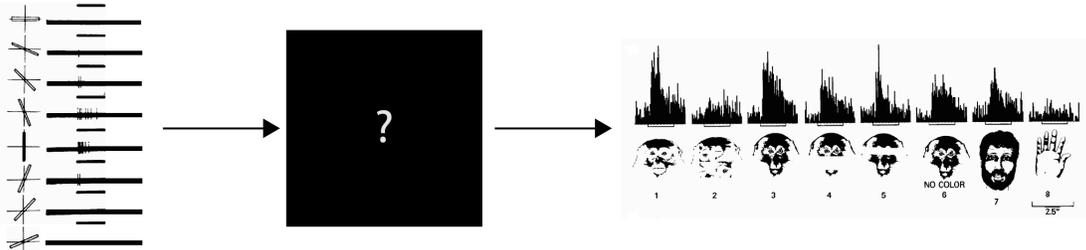
422
423   **5.9.   Divide and conquer**
424
425   To solve the complex task of object recognition, the visual system seems
426   to have adopted a "divide and conquer" strategy. Instead of trying to come up
427   with a single function that will transform lines into complex shapes in one step,
428   the computations underlying pattern recognition are implemented by a cascade
429   of multiple approximately sequential computations. Each of these computations
430   may be deceptively simple and yet the concatenation of such steps can lead to
431   interesting and complex results. As a coarse analogy, consider a factory making
432   cars. There is a long sequence of specialized areas, departments and tasks. One
433   group of workers may be involved in receiving and ordering different parts, others

Table 5.1: Response latencies in different areas in the macaque monkey (from Schmolesky et al 1998).

Area	Mean (ms)	S.D. (ms)
LGNd M layer	33	3.8
LGNd P layer	50	8.7
V1	66	10.7
V2	82	21.1
V4	104	23.4
V3	72	8.6
MT	72	10.3
MST	74	16.1
FEF	75	13

may be specialized in assembling the carburetor, others in painting the exterior. The car is the result of all of these sequential and parallel steps. To understand the entire mechanistic process by which a car is made, we need to dig deeper into each of those specialized sub-steps. To understand the mechanisms orchestrating visual object recognition, we need to inspect neuronal ensembles along the ventral visual stream.

434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449

450 **5.10.   We cannot exhaustive study all possible visual stimuli**
451
452   One of the challenges to investigate the function and preferences of
453   neurons in cortex is that we have a limited amount of recording time for a given
454   neuron. Given current techniques, it is simply impossible to examine the large
455   number of possible combinations of different stimuli that might drive a neuron.
456   Consider a simple scenario where we present a 5x5 image patch where each
457   pixel is either black or white (**Figure 5.6**). There are 2^{25} such stimuli. If we
458   present each stimulus for 100 ms and we do not allow for any intervening time in
459   between stimuli, it would take more than 5 weeks to present all possible
460   combinations. There are many more possibilities if we allow each pixel to have
461   three colors (Red, Green, Blue) with an intensity between 0 and 255. We can
462   typically hold extracellular recordings with single (non-chronic) electrodes for a
463   couple of hours. Recent heroic efforts have managed to track the activity of
464   presumably the same neuron for months (Bondar et al 2009, McMahon et al
465   2014). Yet, even with such chronic electrodes, it is difficult to keep an animal
466   engaged in a visual presentation task for more than a few hours a day. Thus,
467   investigators often recur to a number of astute strategies to decide which stimuli
468   to use to investigate the responses of cortical neurons. These strategies typically
469   involve a combination of: (i) inspiration from previous studies (past behavior of
470   neurons in other studies is a good predictor of how neurons will behave in a new
471   investigation); (ii) intuitions about what might matter for neurons (for example,
472   many investigators have argued that real world objects such as faces should be
473   important); (iii) statistics of natural stimuli (as discussed in Chapter II, it is
474   reasonable to assume that neuronal tuning is sculpted by exposure to images in
475   the natural world); (iv) computational models (to be discussed in more detail in
476   Chapters VII-IX); (v) serendipity (the role of rigorous scrutiny and systematic
477   observation combined with luck should not be underestimated). Combining these
478   approaches, several investigators have began to probe the neural code for visual
479   shapes along the ventral visual cortex.

480   5.11.   **We live in the visual past: response latencies increase along the**
481   **ventral stream**
482
483   Vision seems to be instantaneous. You open your eyes and the world is
484   out there, apparently immediately. Visual processing is indeed very fast. Indeed,
485   we argued in Chapter I that the speed of vision is likely to have conferred critical
486   advantages to the first species with eyes and may well constitute one of the key
487   reasons why evolution led to the expansion of visual capabilities. Yet, the
488   intuition that vision is instantaneous is nothing more than an illusion. It takes time
489   for signals to propagate through the brain. A small fraction of this time has to do
490   with the speed of propagation of signals within a neuron, along dendrites and
491   axons. Yet, the within neuron delays are relatively small. In particular, action
492   potential signals within axons that are insulated by myelin can propagate with
493   speeds of about 100 meters per second. Dendrites tend to be shorter than axons
494   and propagation speeds within dendrites is also quite fast. The main reason why
495   vision is far from instantaneous is the multiple computations and integration steps
496   in each neuron combined with the synaptic hand-off of information from one
497   neuron to the next.
498
499   At each processing stage in the visual system, it is possible to estimate
500   the time it takes for neurons in that area to realize that a flash of light was
501   presented. Response latencies to a stimulus flash within the receptive field of a
502   neuron increase from ~45 ms in the LGN to ~100 ms in inferior temporal cortex
503   (Hung et al, 2005, Schmolesky et al 1998) (**Table 5.1**). There is an increase in
504   the average latency within each area from the retina to the LGN to V1, to V2, to
505   V4, to ITC. This progression of latencies has further reinforced the notion of the
506   ventral processing stream as an approximately hierarchical and sequential
507   architectures. Each additional processing stage along the ventral stream adds an
508   average of ~15 ms of computation time. It should be emphasized that these are
509   only coarse values and there is a lot of neuron-to-neuron variability within each
510   area. For example, an analysis in anesthetized monkeys by Schmolesky and
511   colleagues showed latencies ranging from 30 ms all the way to 70 ms in primary
512   visual cortex. Because of this heterogeneity, the distribution of response
513   latencies overlap and the fastest neurons in a given area (say V2) may fire
514   before the slowest neurons in an earlier area (say V1). The notion of sequential
515   processing is only a coarse approximation. The response latencies constrain the
516   number of computations required to perform computations along the visual
517   hierarchy.
518
519   Because of these latencies, we continuously live in the past in terms of
520   vision. The notion that we only see the past events is particularly evident when
521   we consider distant stars. The light signals that reach the Earth have left those
522   stars a long time ago. Although much less intuitive, the same idea applies to
523   visual processing in the brain. Of course, the time it takes for light to bounce on a
524   given object and reach the retina is negligible, yet signal propagation in the brain
525   takes on the order of a hundred milliseconds as discussed above. In several

Figure 5.6: With current techniques, we cannot exhaustively sample all possible stimuli. Here we consider a 5x5 grid of possible binary images (top) or possible grayscale images (bottom). Even for such simple stimuli, the number of possibilities is immense (see text).
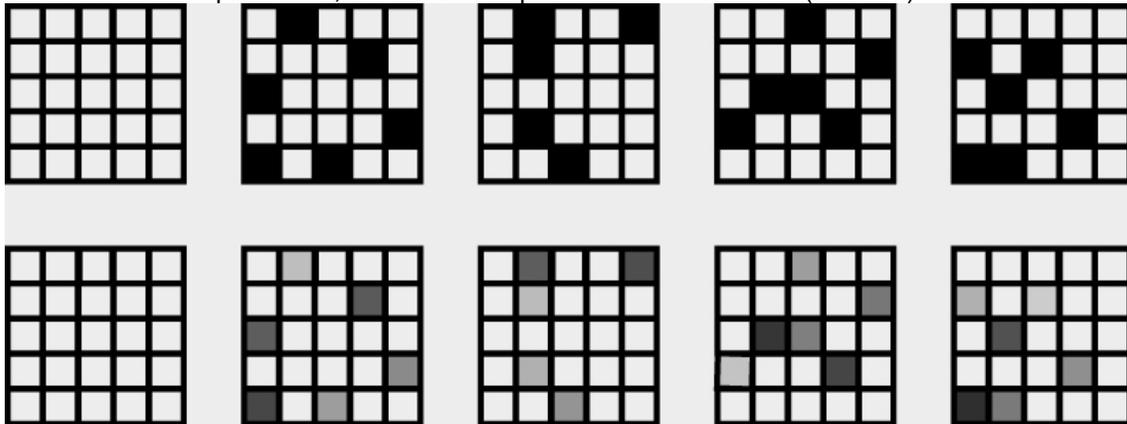
526 cases, through learning, the brain might be able to account for these delays by
527 predicting what will happen next. For example, how is it possible for a ping-pong
528 player to respond to a smash? The ball may be moving at about 50 km/h
529 (apparently, the world record is about 112 km/h) and thus traverses the ~3 m
530 distance in about 200 ms. By the time the opponent has to hit the ball back, his
531 or her visual cortex are processing sensory inputs from the time when the ball
532 was passing the net in the best-case scenario. Not to mention that to orchestrate
533 a movement also takes time (signals need to travel from the decision centers of
534 the brain all the way to the muscles). The only way to play ping-pong and other
535 sports is to use the visual input combined with predictions learnt through
536 experience. Because of these predictions, players not only capitalize on
537 smashing speed but also recur to other strategies such as embedding the ball
538 with spinning effects to confuse the opponent.
539
540 5.12.   **Receptive field sizes increase along the visual hierarchy**
541
542   Concomitant with the prolonged latencies, as we ascend through the
543 visual hierarchy, receptive fields become larger (**Figure 5.7**). Receptive fields
544 range from below one degree in the initial steps (LGN, V1) all the way to several
545 degrees or even in some cases tens of degrees in the highest echelons of cortex
546 (Kobatake & Tanaka 1994, Rolls 1991). Each area has a complete map of the
547 visual field, thus the centers of the receptive fields go from the fovea all the way
548 to the periphery. As discussed for primary visual cortex, within each area, the
549 size of the receptive field increases as we move farther away from the periphery.
550 There is always better resolution in the fovea, across all visual areas. The range
551 of receptive field sizes within an area also increases with the mean receptive field
552 size. The distributions are relatively narrow in primary visual cortex but
553 investigators have described a wide range of receptive field sizes in V4 or inferior

Figure 5.7: Receptive field increases within eccentricity for a given area and receptive field increases along the ventral visual stream at a fixed eccentricity. **a**. Experimental measurements based on neurophysiological recordings in macaque monkeys. B. Schematic rendering of receptive field sizes in areas V1, V2 and V4. Reproduced from Freeman and Simoncelli, 2013.
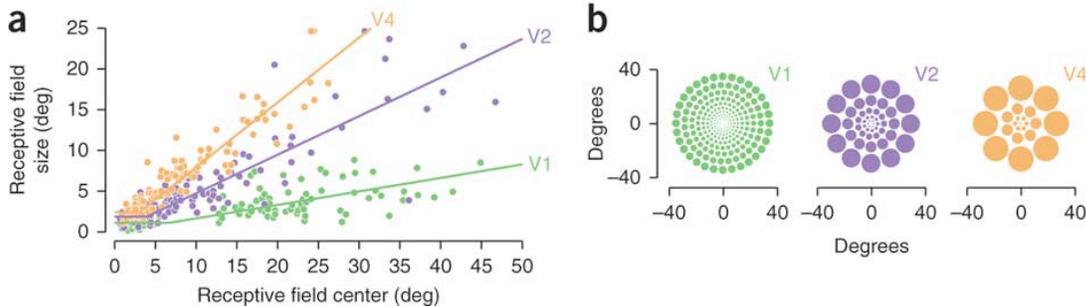
554 temporal cortex. The scaling factor between receptive field size and eccentricity
555 is more pronounced in V4 than in V2 and in V2 compared to V1.
556
557 **5.13.   What do extrastriate neurons prefer?**
558
559   There have been a few systematic parametric studies of the neuronal
560 preferences in areas V2 and V4. These studies have clearly opened the doors to
561 investigate the complex transformations along the ventral visual stream. Despite
562 multiple interesting studies comparing responses in V1, V2 and V4, there isn't yet
563 a clear unified theory of what neurons "prefer" in these higher visual areas. Of
564 course, the term "prefer" is an anthropomorphism. Neurons do not prefer
565 anything. They fire spikes whenever the integration of their inputs exceeds a
566 given threshold. Investigators often speak about neuronal preferences in terms of
567 what types of images will elicit high firing rates.
568
569   The notion that V1 neurons show a preference for orientation tuning is
570 well established, even if this only accounts for part of the variance in V1
571 responses to natural stimuli (Carandini et al 2005). There is significantly less
572 agreement as to the types of shape features that are encoded in V2 and V4.
573 There have been several studies probing responses with stimuli that are more
574 complex than oriented bars and less complex than everyday objects. These
575 stimuli include sinusoidal gratings, hyperbolic gratings, polar gratings, angles
576 formed by intersecting lines, curvatures with different properties, among others
577 (Hegde & Van Essen 2003, Hegde & Van Essen 2007, Kobatake & Tanaka
578 1994, Pasupathy & Connor 2001). Simple stimuli such as Cartesian gratings can
579 certainly drive responses in V2 and V4. As a general rule, neurons in V2 and V4
580 can be driven more strongly by more complex shapes. As discussed above in the
581 context of latency, there is a wide distribution of stimulus preferences in V2 and
582 V4.
583

584
585
586
587
588
589
590
591
592
593
594

Figure 5.8: **V2 cells can represent lines that do not exist except in the eyes of the beholder.** The figure shows the Kanizsa triangle visual illusion and a schematic rendering of neurophysiological recordings from 4 neurons: two retinal ganglion cells neurons (RGC) and two V2 neurons. When the receptive fields (gray dotted circles) encompass locations that have a real contour (A), both RGC and V2 cells fire vigorously. In contrast, when the receptive fields encompass an illusory contour (B), the V2 cell fires vigorously but the RGC cell only fires a few baseline spikes.
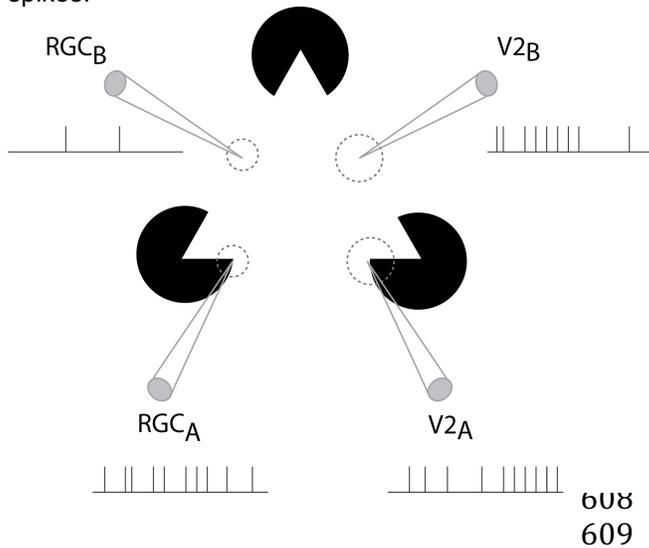
Perhaps one of the challenges is that investigators seek an explanation of neural coding preferences in terms of colloquial English expressions such as orientation, curvature, etc. An attractive idea that is gaining momentum is the notion that neurons in these higher visual areas filter the inputs from previous stages to produce complex tuning functions that defy language-based descriptions. A neuron may be particularly activated by a patch representing complex shapes and textures that is not simply defined as an angle or a convex curve. Ultimately, the language of science is mathematics, not English or Esperanto. Neuronal tuning properties do not have to map in any direct way to a short language-based description.

608
609

610   **5.14.   Brains construct an interpretation of the world: the case of illusory**
611   **contours**
612
613   Another pervasive illusion is that our senses contain a veridical
614   representation of exactly what is out there in the world. This notion can be readily
615   debunked through the study of visual illusions (Chapter I). Let us revisit the
616   Kanizsa triangle (Chapter 1, **Figure 5.8**) where we have the strong illusion of
617   perceiving an equilateral triangle in the midst of the three pacman icons. The
618   sides of the triangle near the vertices are composed of real black/white contours.
619   However, the center of each side is composed of a line that does not really exist.
620   These lines represent illusory contours, edges created without any change in
621   luminance. It is relatively easy to "trick the eye". Except that the eye is typically
622   *not* tricked in most visual illusions. Visual illusions represent situations where our
623   brains construct an interpretation of the image that is different from the pixel level
624   content. In most such illusions, retinal ganglions cells do follow the pixel level
625   content in the image relatively well. If we record the activity of a retinal ganglion
626   cell whose receptive field is right in the center of the illusory contour, nothing
627   would happen upon flashing the Kanizsa figure. In other words, the activity of
628   retinal ganglion cells does not correlate with our perception. But if the retina does
629   not reflect perception, then who does? It seems reasonable to conjecture that

630 there must be neurons somewhere that explicitly represent the contents of our
631 perception, in this case the illusory contours (this is a critical postulate that we
632 will discuss again in more depth when we take up the question of the neuronal
633 correlates of consciousness in Chapter VII).
634
635   Indeed, neurons in area V2 respond vigorously to illusory contours
636 (Figure 5.8). These V2 neurons respond almost equally well to an illusory line or
637 to a real line (Lee 2003, Lee & Nguyen 2001, von der Heydt et al 1984). The
638 responses to illusory contours are remarkable because there is no contrast
639 change within the neuron's receptive field. Hence, these responses must indicate
640 a form of context modulation that is consistent with the subjective interpretation
641 of borders. There are also neurons in V1 that respond to illusory contours but
642 there are more such neurons in V2.
643
644 **5.15.   A colorful V4**
645
646   Neurons in area V4 are particularly sensitive to stimulus color (Zeki
647 1983). Neurons in area V4 demonstrate sensitivity to color properties that are
648 more complex than those observed in earlier areas such as LGN parvocellular
649 cells or V1 blobs. Neurons in V4 have been implicated in the phenomenon of
650 color constancy whereby an object's color is relatively insensitive to large
651 changes in the illumination, in contrast to the responses earlier in the visual
652 system.
653
654 **5.16.   Attentional modulation**
655
656   As discussed for V1, neurons along the ventral visual cortex receive
657 massive top-down signals in addition to their bottom-up inputs (Markov et al
658 2012). Presumably through these top-down signaling mechanisms, the activity of
659 neurons along ventral visual cortex can be strongly modulated by signals beyond
660 the specific content of their receptive fields including spatial context and higher
661 level cognitive influences such as task goals.
662
663   A prime example of this type of modulation involves spatial attention
664 (Desimone & Duncan 1995, Reynolds & Chelazzi 2004). Importantly, spatial
665 attention effects can be demonstrated outside of the fixation focus. That is, a
666 subject can be looking at one place and paying attention to another place, a
667 phenomenon known as *covert* attention (as opposed to overt attention which is
668 the more common scenario where attention is allocated to the fixated area).
669 Through a series of astute training paradigms, investigators have been able to
670 train animals to deploy covert attention, thus enabling them to investigate the
671 consequences of spatial attention on neurons with receptive fields outside the
672 fovea. Neurons typically show an enhancement in the responses when their
673 receptive field is within the locus of attention. The magnitude of this attentional
674 effect follows the reverse hierarchical order, being significantly stronger in area
675 V4 compared to area V1.
676

677 **References**
678
679

680 Bondar I, Leopold D, Richmond B, Victor J, Logothesis N. 2009. Long-term stability of
681   visual pattern selective responses of monkey temporal lobe neurons *PloS one*
682   9
683 Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, et al. 2005. Do we know what
684   the early visual system does? *Journal of Neuroscience* 25: 10577-97
685 Desimone R, Duncan J. 1995. Neural mechanisms of selective visual attention.
686   *Annual Review of Neuroscience* 18: 193-222
687 Hegde J, Van Essen DC. 2003. Strategies of shape representation in macaque visual
688   area V2. *Vis Neurosci* 20: 313-28
689 Hegde J, Van Essen DC. 2007. A comparative study of shape representation in
690   macaque visual areas v2 and v4. *Cerebral cortex* 17: 1100-16
691 Hubel D, Wiesel T. 1959. Receptive fields of single neurons in the cat's striate cortex.
692   *Journal of Physiology (London)* 148: 574-91
693 Hubel DH, Wiesel TN. 1968. Receptive fields and functional architecture of monkey
694   striate cortex. *The Journal of physiology* 195: 215-43
695 Hung CP, Kreiman G, Poggio T, DiCarlo JJ. 2005. Fast Read-out of Object Identity
696   from Macaque Inferior Temporal Cortex. *Science* 310: 863-66
697 Kobatake E, Tanaka K. 1994. Neuronal selectivities to complex object features in the
698   ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71:
699   856-67
700 Lee TS. 2003. Computations in the early visual cortex. *J Physiol Paris* 97: 121-39
701 Lee TS, Nguyen M. 2001. Dynamics of subjective contour formation in the early
702   visual cortex. *Proceedings of the National Academy of Sciences of the United*
703   *States of America* 98: 1907-11
704 Markov NT, Ercsey-Ravasz MM, Ribeiro Gomes AR, Lamy C, Magrou L, et al. 2012. A
705   Weighted and Directed Interareal Connectivity Matrix for Macaque Cerebral
706   Cortex. *Cerebral cortex*
707 McMahon DB, Jones AP, Bondar IV, Leopold DA. 2014. Face-selective neurons
708   maintain consistent visual responses across months. *Proceedings of the*
709   *National Academy of Sciences of the United States of America* 111: 8251-6
710 Pasupathy A, Connor CE. 2001. Shape representation in area V4: position-specific
711   tuning for boundary conformation. *J Neurophysiol* 86: 2505-19
712 Reynolds JH, Chelazzi L. 2004. Attentional modulation of visual processing. *Annu Rev*
713   *Neurosci* 27: 611-47
714 Rolls E. 1991. Neural organization of higher visual functions. *Current opinion in*
715   *neurobiology* 1: 274-78
716 Schmolesky M, Wang Y, Hanes D, Thompson K, Leutgeb S, et al. 1998. Signal timing
717   across the macaque visual system. *Journal of Neurophysiology* 79: 3272-78
718 von der Heydt R, Peterhans E, Baumgartner G. 1984. Illusory contours and cortical
719   neuron responses. *Science* 224: 1260-62
720 Zeki S. 1983. Color coding in the cerebral cortex - The reaction of cells in monkey
721   visual cortex to wavelengths and colors. *Neuroscience* 9: 741-65
722