Chapter 6

Routines

Up to this point you know a lot about the brain’s programs. They sample the world in a cortical instant, that is 100 to 300 milliseconds duration, and use that information to compute the state that is used to decide what to do next. Furthermore the selection of the subsequent options is managed by the Basal Ganglia. And the alternatives that do come up are rated by the dopaminergic system. A lot of details are missing but in broad outline you have a initial picture of the brain’s system for managing the programs that govern behavior. However one of the most important issues that has been finessed is the computation at the interface of the all-important state needed to guide programs. While we touched on the mechanism for computing state, the dynamics driving content addressable memory, an actual discussion of the how its content was created was postponed. The reason is that the body which the brain finds itself in plays a huge role in this context. The body has been spectacularly designed by evolution to compute just those things that it needs to direct behavior. Basically, the body can be thought of as a special purpose computer and the computation that it does, the brain does not have to repeat. Moreover, given that the body has done a large part of the necessary computation, the computation left over for the brain to do is drastically simplified. In this chapter we explore these relationships.

6.1 Special purpose computation

The idea that the body computes things can be subtle, so to develop it we will start by revisiting the issues related to the cost of computation. Worrying
whether a computation can be done at all is the subject of computability. You have already seen that the centerpiece of the answer to computability questions is the Turing Machine; something is computable if a TM can do it. But another important question concerns the cost of computing something that we know is computable in the first place. Here we have to be a little careful when interpreting classical computer science results in terms of brain computation.

The classical result is that, when cost is the issue and the size of the data it must handle is not constrained, it’s the algorithm that is paramount. When you have an algorithm that has to do something to \( N \) data points, the performance of the algorithm can be described in terms of some function of \( N \), denoted \( O(f(N)) \), independent of the machine that implements the algorithm. The key assumptions in this classical result are that 1) the algorithm cost is reported as the worst case and 2) The input is allowed to grow arbitrarily. But the first assumption is very conservative: if it took only \( N \) operations on almost all the data sets that occurred, but \( O(N^3) \) on a just few hard cases, it is still reported as \( O(N^3) \). As for the second assumption, the mature brain is of a fixed size, and thus the inputs cannot grow arbitrarily; biology works with fixed size inputs. For example the retina contains about \( 10^8 \) light sensitive cells and the other sensory input and motor outputs have similar bounds. When the input is bounded in this way it pays to make the basic computer faster. Special purpose computers can be very fast indeed.

Additionally, in brain computation, the classical results are too conservative in another way. The brain is happy to use approximate algorithms that get the answers right some of the time (as any SAT taker knows!) And in everyday life if one way of doing something fails there is usually another way.

All these differences between classical silicon accounting conventions and the brain’s special tricks allow the brain to function in real time with its slow circuitry, but perhaps the most stunning and unappreciated component is the body itself. Computation need not be only done by a standard computer. In a wonderful example from Fischler and Eschlager, consider finding the shortest path between two selected points in a graph such as shown in Figure 6.1. The standard expensive algorithm would start at one end and check all the path lengths to the nearby nodal points, and then repeat the process from each of these. But what about building a physical device that represents the particular graph exactly in terms of having string connecting balls representing the nodes? Now to find the shortest path between any two nodes, you can just pick them up and pull them apart until the string is taut.
6.2. ROUTINES DEPEND ON VISUO-MOTOR CONTEXT

Figure 6.1: Finding the shortest path in a graph is easy if a physical embodiment of the graph is available. Upper A graph with links between edges constructed from balls and string. Lower Pulling the two ends apart results in the shortest path being horizontal.

The nodes on the taut path are the answer!

The bodies of animals are like Fischler and Esclager’s special purpose path finder in that they too are optimized for just the kinds of problems that come up in getting about in the world. The sensory motor coordination that they do would be extremely expensive if done in a general purpose way, but that is not what happens. Instead evolution eschews general purpose computations for blazing speed in a special purpose computer simply known as the body.

6.2 Routines depend on visuo-motor context

Because the ultimate behaviors are motor acts and these are carried out a particular body, the codes that the brain uses only have to have the fidelity to direct that body. This point is so important that we will try and emphasize it with what first might seem like a digression: The use of bodies by insects, in this case ants.
The example of insect bodies: weaver ants  Ants are the most remarkable creatures (and almost the most numerous: two-thirds of the biomass is comprised of ants and termites). A tremendous amount of what we know about them comes from the work of two researchers Bert Hölldobler and Edward Wilson. Let’s just focus on one genus of ant, the weaver ants. These ants build nests out of large tropical leaves by bending them over and gluing the result. The problem of bending the leaves is complicated by the fact that they are much larger than a single ant. What the ants do is they form an ant chain to grab the end of the leaf and slowly crank it over. This leaves the problem of gluing. The ants’ own larvae make glue. So the ants have to pick the larvae up and maneuver them back and forth, squeezing them to get the glue in the right places.

The amazing behavior of ants is easy to describe as a program. Their internal gluing program is the equivalent of:

```plaintext
find larva
pick up larva
move to the right spot on leaf
squeeze larva
```

Ants use vision but most of them also have an arsenal of chemicals that they can use to signal other ants and mark locations. These instructions work because of the ant’s body. They are designed for it and depend on its construction and attributes. They also are functional. Consider the use of larvae in our tentative model program. What is a larva to an ant? We don’t know but we can be sure that it is a very minimal functional test that works because on a leaf there aren’t any competing alternatives. To drive this point home, consider that horseshoe crabs, an evolutionary relic that is shaped rather dome-like, will mate with an overturned bowl. It’s not that they are not fussy; its rather that their test for a mate is rudimentary and the bowl passes muster. In the same way the larva test can be very spartan. In fact the tests can be different for each of the steps. For example “squeeze larva” can be actually just “squeeze” which works fine at the point where a larva is being held over the right spot on the leaf. We also know that the sequence of instructions is very rigidly linked.

The main point is that it’s the ant’s body that makes this minimalism work. Its biomechanics has been refined by evolution to work in its ecological niche and the constraints that are needed to get the job of existing and
6.2. ROUTINES DEPEND ON VISUO-MOTOR CONTEXT

procreating done are hard coded - literally - in the degrees of freedom of the ant’s exoskeleton.

**Human visuo-motor routines exploit context** Visuo-motor computations have to be efficient since their result often has to be computable in a few hundreds of milliseconds. Thus, for the most part, the computations are task-dependent tests since those can use prior information to simplify the computations. Suppose you are looking for your favorite cup that happens to have an unusual color. Then rather than any elaborate parsing of the visual surround, you can simply look for a place that has that color. Motor routines need to be efficient for the same reason that visual routines are. Since they are also goal directed, they can make extensive use of expectations that can be quickly tested. If you are searching for the same cup on a nearby table in the dark, you can use the remembered shape of the cup to quickly and efficiently interpret haptic data from your fingers and palm.

Consider another example of walking down a sidewalk with instructions to stay on the sidewalk, avoid objects of a blue color and pick up objects of a purple color. As table shows each of these tasks can be handled separately and each of the corresponding vision and motor routines is very simple.

<table>
<thead>
<tr>
<th>Task</th>
<th>Visual Routine</th>
<th>Motor Routine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stay on sidewalk</td>
<td>See if sidewalk edge crosses midline</td>
<td>if so turn away</td>
</tr>
<tr>
<td>Pick up purple objects</td>
<td>Find closest purple thing</td>
<td>Head toward it</td>
</tr>
<tr>
<td>Avoid blue objects</td>
<td>Find closest blue object</td>
<td>If its heading is unchanging, veer</td>
</tr>
</tbody>
</table>

Table 6.1: Different tasks have specialized uses of vision as well as specialized motor responses.

Figure 6.2 shows three different visual tests used in three separate tasks on the same image. The bottom line is that human visuo-motor behaviors are directed in a compact purposeful way. Like the ants, they can also be solved by extracting just the information needed to orchestrate the immediate sub-task. The huge difference is that the human brain is programmable and can learn an enormous library of routines from experience.
Figure 6.2: Walking down a sidewalk in a virtual world. Three different tasks can be solved differently by special purpose visual processing. Finding a litter box that is known to be purple can be handled by using a purple color template. Staying on the sidewalk can be handled by finding the edge of the sidewalk which is of a known range of orientations with respect to the viewer. Avoiding obstacles can be handled by using a depth map to extract nearby obstacles that are impeding the walker's trajectory.
6.3 Human embodiment overview

At this point we have covered two major points. The first is that the human body has been especially designed by evolution to solve data acquisition and manipulation problems. The second is that the solutions to these problems are very compact owing to the fact that they are purposive. But a further important point is that these do not have to be re-solved by the forebrain. Instead the forebrain can access and direct the form of these solutions using its own very compact, abstract representations that are tailored to the body’s design. Furthermore, as we will elaborate, the vision and motor system solutions have essential cortical features in common.

Dynamic reference frames  Any representation that involves geometrical data has to use some kind of reference frame to describe the data. Huge debates have ranged over the coordinate system used in vision: Is it head based, eye based or otherwise? The fixation system shows that it must be dynamic. Depending on the task at hand, it can be any of these. Imagine using a screwdriver to drive a screw into hardwood. The natural reference frame for the task is the screw head where a pure torque is required. Thus the gaze is needed there and all the muscles in the body are constrained by this purpose: to provide a torque at a site remote to the body. To keep this idea of a remote frame of reference in mind, consider how the visual system codes for depth. In the early stages of the cortex, the neurons are sensitive to small differences or disparities between the images in the left and right eyes. Roughly speaking, neurons that have disparity-sensitive receptive fields code for zero, negative and positive disparities. However consider zero disparity. This measurement is produced when the two images match and this is at the fixation point, a point not in the body at all, as depicted in Fig. 6.3.

Like visual routines, motor routines also use dynamic frames of reference. The screw driving example used for vision also applies for motor control. The multitude of muscles work together to apply a pure torque at the screwdriver end. Another example is balance. For an upright stance the motor system must make sure the center of gravity is over the base of the two feet. But the center of gravity is a dynamic point that moves with postural changes and on these excursions can easily be outside of the body. Nonetheless, inside or out, the control system must refer its computations to this point.

Consider the act of grasping a drinking tumbler. The best way to do this would be to orient the hand’s grip so that the axis of the grasp aligns
Figure 6.3: Most of human hand eye coordination can be directed by discrete gaze and hand targets. The hands and eyes use egocentric frames of reference (Red frames), whereas objects in the world have their own natural allocentric coordinate systems (Green).

with the axis of the tumbler. But of course the hand’s orientation typically does not start out that way at the initiation of the grasping action. The problem is to align the two frames of reference. The tumbler frame is termed an *allocentric* frame as it is outside of the body. Conversely the hand fame is termed an *egocentric frame*. The visual system also uses an egocentric frame, but as just discussed, it is not attached to the body either but floats in space at the point of fixation.

**Cortical organization anticipates primate embodiment** Visually guided grasping is a nice example, because its our hand eye coordination ability that really sets primates apart from the rest of the animals. Unsurprisingly, in our corticies, both the central area of the visual field and one the central areas of the body, the hands, have expanded representations. The small one degree of central vision that has over 100 times the resolution of the periphery is known as the fovea. Unsurprisingly the hand area also is hugely exaggerated with respect to the rest of the body (excluding the face area) in motor cortex area M1. There is not really a name for this, but it’s a kind of ‘movement fovea.’ For our discussion, lets take liberties and call it a ‘movea.’ At this point, we can compare the visual and motor corticies and see that hey have a parallel organization. Vision has two phases 1) orienting the gaze and 2)
extracting information from the gaze. It uses peripheral egocentric information to orient the gaze located in dorsal cortex and then it uses the temporal cortex to extract allocentric information from the fixated area. Think of this as a ‘point and read’ strategy. The readout of the fixation system locates the tumbler with respect to current position of the body and the readout of the foveal routines supplies necessary allocentric information. This information is connected to the motor side via the abstract areas of the corticies. In this connection pattern, the motor corticies are similarly organized. The egocentric information computes how to change the body’s posture to grasp the tumbler in the large and the allocentric information sent to the ‘movea’ determines the fine points of the grasp. This organization is summarized in Fig. 6.4.

Figure 6.4: The organization of the cortex in human hand eye coordination shows similarities that reflect the overlying organization of the body. Both the visual and motor areas have exaggerated central representations that are used to extract and coordinate allocentric information.

**Gaze stabilization** The visual system has six separate systems to stabilize gaze. This of course is an indication of just how important it is to achieve gaze stabilization but the third consequence is that the visual computations can be simplified as they do not have the burden of dealing with gaze instability. Despite the fact that the human may be moving and the object of interest may be independently moving as well, the algorithms used to analyze that object can assume that it remains in a fixed position near the fovea.
Human vision uses fixations that have an average duration of 200 to 300 milliseconds. The fixational system brings home the key role of embodiment in behavior. Although the phenomenal experience of vision may be of a seamless three-dimensional surround, the system that creates that experience is discrete. Furthermore as humans are binocular and make heavy use of manipulation in their behaviors, they spend most of their time fixating objects in the near distance. That is the centers of gaze of each of the eyes meet at a point in three dimensional space and, to a first approximation, rest on that point for an average of three hundred milliseconds.

Compliance in the motor system The adult human motor system is comprised of an extensive musculo-skeletal system that consists of over two hundred bones and somewhere between 650 and 850 muscles. One of the most important properties of this system is the passive energy that can be stored in muscles. This spring-like system has at least two important properties: 1) it can lead to very economical operation in locomotion, and 2) it can be used in passive compliant collision strategies that save the system from damage. Moreover it can be driven by a discrete strategy whereby set points for the spring-muscle system are communicated at a low bandwidth. Simulations show that reasonable motions can be obtained with sampling intervals approximating the fixation interval.

There are many examples that could be mentioned to illustrate how the computation done by the motor system makes the job of motor routines easier, but one of the most obvious is the extensive use of passive compliance in grasping. If the motor system was forced to rely on feedback in the way that standard robot systems do them the grasping strategies would have to be far more delicate. Instead, grasp planning can be far easier as the passive conformation of the multi-fingered hand provides great tolerances in successful grasps. Figure 6.5 shows a virtual figure demonstrating that grasping is straightforward.

The WHAT/WHERE dichotomy Computation can be greatly simplified by recognizing that the properties of an object can be divided into those that are associated with the object itself and those that are associated with the object’s relation to other objects. The former are related to the object’s identity or WHAT the object is and the latter are related to the object’s position or WHERE it is. Tackling these computations separately greatly
Figure 6.5: The natural springiness of the musculo-skeletal system coupled with the use of discrete targets can lead to very simple grasping strategies. Here a model human simply has to aim for the center of a cylinder and open and close the hands. The natural properties of the hand do the rest. A) The egocentric computation aims the hand at the tumbler. B) The allocentric computation orients the grasp.
Figure 6.6: An illustration of the fundamental dichotomy of routines into WHAT and WHERE using the feature of color.  
A In a WHERE problem the target object’s colors are known but the position needs to be determined. This can be done by computing where the target colors are clustered in the scene. In the upper left image, a dense cluster is a probable location of the colored object.  
B. Two examples of a WHAT problem being solved. In a WHAT problem the location is known and the colors from that location - in these cases the entire image - are compared to colors associated with internal models of objects. The insets show the top three models matched.

simplifies the costs involved. Figure 6.6 shows an example using color features.

You will have recognized that the WHAT corresponds to the use of allocentric features gained at the point of fixation and the WHERE corresponds to egocentric features measured using the full field of view. Arguments break out among scientists as to whether these distinctions can be so orthogonal, but, appealing to the central theme of computational abstraction, the answer depends on the particular level. One level down leads to the CAM memory models of cortex and there the reas would interact. However the algorithmic behavior can make the result appear as if the distinction is being made. To
locate a pattern, one can fix the pattern’s cortical code and let the dynamics settle accordingly. Similarly to identify a pattern, fixating it has the effect of fixing the input code so that the dynamics can settle to a labeled internal model.

6.4 Visual routines

When we think of movements, it seems obvious that the motor program makes choices and communicates those choices to the spinal chord. But it would be a little strange to think of the sensory systems, in particular the visual system as working in the same way. Or would it? In fact evidence is mounting that the visual system may work in just that way. The brain runs a program to do a certain task. That program needs information about the visual world in order to proceed. So visual tests are commissioned to get the requisite information.

The view that vision works this way is tough to swallow because our conscious experience suggests that it works in a different way. The visual world seems constantly present in a rich tableau that we can inspect at our leisure. That this is not the case was given dramatic voice in experiments done by Dan Simons. These experiments were specifically directed at trying to pin down just what we notice about our surroundings in everyday situations. A student would approach a professor on the Cornell campus and ask for directions. Cornell has an extensive campus so it took time to get the directions right. While the professor was answering, two rude students carrying a door would barge between the professor and his interlocutor. Unbeknown to the professor, the three students were in cahoots: As the two passed through, the rear door carrier and the interlocutor would change places. They had practiced this maneuver and carried it off smoothly. The surprising result was that 50% of the professors did not notice anything unusual by carried on talking to the new arrival as if nothing had happened. Since that time Simons has explored many different variants of this situation. In one setting an customer checking in at a hotel sees the clerk duck below the counter for a moment only to be talking to a new person who pops up in his place. In another subjects watch movies where the contents of the scene are changed between cuts. But the results are always the same. Large numbers of people are not aware of any changes.

We need only be surprised at these results if we are lured into adopting
the expectations of conscious experience: We think we see everything so if something changes, gosh darn it, we should notice. However if we adopt the program-centered view then the visual processing done is at the behest of the program for very specific purposes. The hotel guest only has to be sure that she is having a conversation with a clerk that will lead to being checked in. The basics are that a person is swiping the credit card and coming up with the room key. A moving thing that passes the person test as well as the card and key particulars will be good enough to do the job. Basically the strategy is the same as that used by the insects. The difference is that the tests can be composed as they are not as directly tied to the motor system.

Cortical evidence for routines If the brain is working this way then given what we know about the cortical moment, the time for these alleged tests is pinned down; they should be carried out within the periods where gaze is fixed. Some beautiful experiments by Peiter Rolfsema’s group in Amsterdam suggest just that. They trained monkeys to do a visual line tracing task depicted in Fig. 6.7. The monkeys had to hold their gaze fixed and determine which of two lines was connected to a central point. Both lines terminated near the point so it was not too easy. Next the monkeys had to look at the end of the line that passed the test. The rationale behind the experiment is that the monkey in computing the end point would trace the locus of the correct line. Since the neurons in the cortex form a visual map, a neuron on the line’s locus should be activated by the tracing process. The trick in the experimental manipulation is that it takes the monkey a few hundred milliseconds to program the gaze change so that during that time gaze is fixed and the visual world projection onto the cortex is stable. Thus they could record the output of a neuron along the tracing path with the hope that they would see a tracing-related signal.

To the experimenter’s delight, at 90 milliseconds after the trigger signal, the neuron about halfway along the path fired, suggesting that the line was indeed being mentally traced. Next the experimenter’s introduced a more complicated variant of this protocol. The ends of the lines and a nearby central dot had colors. The monkey now had to look at the end of the line that terminated with the same color as the central dot. The hypothesis of course was that since the internal program now has three steps namely:

determine the line with the matching color;
trace that line to the other end;
6.4. VISUAL ROUTINES

Figure 6.7: An experiment with primates provides evidence for visual routines. The top panel shows an experiment where a monkey has to stare a point and then when a pair of lines appears, look at the end of the line that is attached to the point of fixation. The fact that programming the eye movement takes 250-300 ms means that the eyes are still for the interim. Recordings from edge cells along the path show elevated firing beginning at 130 ms suggesting that the monkey’s brain is using a visual routine to trace the path. In the bottom two panels the task is made harder and the result is that tracing evidence starts later (see text). [Permission Pending]

look at the other end;

then the line tracing operation would be done slightly later as it would have to await the result of the color matching step. Sure enough, in this case the neurons along the path fire about 60 milliseconds later. This work is some of the most convincing evidence yet on the use of specific tests in the course of visual problem solving. In this case the task has been arranged so that the best way to solve it is to do specific tests in a specific order. But the thought is that in general, it may be that visual processing is broken down into specific tests like these.
Psychophysical evidence for routines  The monkey experiments are a tour de force but still, it would be good to have corroborative evidence that humans are doing the same thing. It turns out that there are a number of highly suggestive psychophysical experiments all pointing in the same direction.

A computational experiment by Salgin tested the use of a looming routine to stay a constant distance behind a car in a virtual environment. The key trick was to transform the image from cartesian $x$ and $y$ coordinates into and equivalent $\Theta$ and $\log r$ system. The latter system is used by the cortex and has some nice properties, one being that an expansion or contraction of the image becomes a shift which is very easy to detect algorithmically. Figure 6.8 shows how it works. The outline of a braking truck is just shifted to the right because as it comes closer, its radius gets larger. A visual routine can easily measure the shift.

In introducing model visual routines we used the example of multitasking in human walking. Thus Fig. 6.2 showed the computations used by a model of the humans’ behavior. But using immersive virtual environments, actual humans can carry out identical scenarios. It turns out that they have different routines for picking up objects and avoiding them and reveal this by using different fixation patterns in each of the two cases. Figure 6.9 shows the fixation points as dots in each case.

Our final set of examples were conducted by Treisch et al. when he was at the University of Rochester. The Computer Science department there has an extensive capability in virtual reality that includes the ability to create artificial sensations of force when picking up a virtual object using a dual Phantom force producing system.

In Treisch’s experiment, subjects picked up virtual blocks and them on one of two conveyor belts whereupon they were whisked away. The subjects sorted the blocks by a size feature. The subjects were told that the software running the experiment was buggy and to report any changes. Three separate sorting conditions were used that varied according to different initial instructions. The differences were chosen to with the view to changing the subject’s internal program. Here they are:

1. “Pick up the blocks in front to back order and put them on the closest conveyor belt.”

2. “Pick up the tall bricks and put them on the conveyor belt. Then pick up the small bricks and put them on the closer conveyor belt.”
Figure 6.8: A model of the human ability to follow a car tested in a virtual environment. The image is transformed into a log-polar representation whereby the coordinates $x, y$ become $\theta, \log r$ using the relationships $\theta = \arctan(y/x)$ and $r = \sqrt{x^2 + y^2}$. In the transformed coordinates the contour of the followed car simply shifts to the left or right depending on whether it is expanding or contracting respectively. This relationship can be easily detected by cortical circuitry in principle.
Figure 6.9: An experiment where subjects walk through a three dimensional virtual environment. Subjects have to approach some targets and avoid others. even though the targets have identical shapes, the subjects’ gaze is used differently in the two cases. When approaching a target, the gaze is directed towards the center of the target. However when avoiding a target, gaze is directed towards one of the edges. In the latter case subjects are keeping their distance and by implication using a very different algorithm than when approaching.
6.4. VISUAL ROUTINES

3. “Pick up the tall bricks first and put them on the closer conveyor belt. Then, pick up the small bricks and put them on the distant conveyor belt.”

On 10% of the trials the size of the block was changed en route. Since this is a virtual reality environment this can be done seamlessly. If you think about these instructions for a moment you will no doubt guess the hypothesis. The importance of size is varying in the different instructions. In the first case size is irrelevant; in the second, size is important at the time of pick up; and in the third case it is important both at the time of pick up and put down. So if subjects are using an internal program that tests for size, they should notice more of the changes when size was relevant. This is exactly what happened. The average of two independent measures showed that subjects noticed about 10% of the changes in case 1; 25% of the changes in case 2; and 50% of the changes in case 3.

This paradigm was expanded by Droll and Hayhoe who used two separate cues to signal the pick up and put down action. For example subjects might be cued to pick up a red block form a choice of red and blue and then at the conveyor belts they would be cued to put red on the left belt. Subjects are instructed to throw any changed blocks that they notice in a virtual trash bin. One of the virtues of this design is that it provides information on the subjects internal state for missed changes. Suppose a red block turns to blue and the change is unnoticed. Will subjects put the block on the conveyor belt indicated by the red cue, or will they put it on the belt indicated by the blue cue? In the first case they registered the color as red and did not repeat the check. In the second case they apparently would have read the put down cue but have forgotten the color at pick up. It turns out that subjects overwhelmingly sort missed changes by the first cue color. These results fit in with the general theme that vision is the result of explicit tests in the agenda of an ongoing program. From this perspective it all makes sense; any program should not keep testing for color because the way the world is organized is that objects do change color. To keep checking would be a waste of effort. But you can see that we are back to the ants. We share the enormous efficiency of just probing the parts of the world that we need to drive behaviors.

If we believe it, this view is shocking, but perhaps no more shocking than the revelation that the stuff of the world is made up of atoms or that atoms are mostly space. But we still need a computational account for the
Figure 6.10: Four frames from the moment by moment eye fixations recorded as a subject places a block on a conveyor belt in a virtual environment. The red circles show the position of the fingers. Even though fixating the block that has changed color, the subject fails to notice the color change. The color is registered by the retina, but apparently not monitored by the observer.
phenomenology of conscious experience. This will be developed more later, but for the moment we will be content with the classical refrigerator light analogy. Think for a moment about the inside of your refrigerator. Is it light or dark in there? Although you might know that the light is switched off when the door closes, you’ve only ever seen its state with the door open. In that case the light is on. In fact every time you check the light is on! This is a good analogy to the mechanisms that drive conscious visual perception. Anytime we think of a way to check our perceptions the test always comes back in a way favorable to the way the world actually is. The light is always on.

6.5 Embodied motor programs

The strategy of coding just the movements that are needed for behavior is repeated and refined as we move up the complexity ladder in the animal kingdom. Emilio Bizzi and Sandro Mussa-Ivaldi have studied frog reflexes extensively and what they found coded in the frog’s spinal cord are movement primitives. In the frog’s spinal cord the neurons are encoding whole time-dependent movement patterns which can be used in combination to produce variations in a basic movement. If a frog wants to scratch itself, perhaps to get rid of a parasite on its back, the location of the scratch site can be adjusted by changing the ratio of basic movements in the combination. The researchers found this out by varying the point of electrical stimulation in the spinal cord and observing that the scratch site changed systematically. If it helps, think of (musical) chords in the spinal cord to remember this idea. A candidate for the world’s most brilliant and prolific composer is Wolfgang Motzart. If you attempted to capture all of his music by digitizing the concert sounds, it would take an enormous amount of space, but if you instead saved the sheet music, all of his compositions would fit on one compact disc. Crudely put, the forebrain has the sheet music but the spinal cord, combined with the body, makes the music. For reflexes such as scratching, the spinal cord can play its own tune.

In mammals like us, movement reflexes are even more complex. Early experiments on cats showed that a cat has highly sophisticated control systems in its spinal cord. If most of the brain of a cat is removed, sparing the spinal chord and brainstem, that cat can walk, with a little support help on a treadmill. Thus the basic coordination of the gait and support of the
Figure 6.11: The human motor system has a huge array of muscles that can be controlled simply [Permission pending].

weight of the body is all coded in the spinal chord. Even more amazing is a small obstacle is placed in the path of a foot, the moment that foot detects the obstacle, it is retracted and lifted in one smooth motion so as to clear the obstacle and return to the treadmill base. As a runner who has inadvertently tripped many times, I have always been grateful for these reflexes, which make balance-restoring movements that I would not have time to think about and consciously correct. But again not to loose track of the main point: From the brain’s vantage point the body, here represented by the musculo-skeletal system and its controlling reflexes, already has a huge variety of such reflexes that evolution has tailored to work in the world. This gives the brain so much less to do. The brain is like the captain of a 1940s steamer calling down to the engine room for speed changes. A huge number of stokers and overseers implement those changes, but these details are hidden from the captain’s view. In the same way the enormous complexity coded in the spinal chord frees the brain to run more strategic programs. Metaphors can carry intuitions but it turns out that we can make these ideas much more precise by using computational models. Chapter three introduced the idea of basis functions for visual codes but this mathematical trick can be used for any function, including the time dependent functions used to generate
movements, like the ones discussed here. We will take up this point shortly by describing examples that model the human motor system, but first we need to revisit the basic equations of motion for a physical system to remind ourselves of just how complicated a massive multi-jointed system can be.

6.6 Motor computation basics

Whether human or robot, all physical systems have to obey Newton’s laws of motion and for massive systems with multiple degrees of freedom the associated equations can be very complex and not easily solved. We will briefly summarize the main issues. Movement generation involves both kinematics, the geometric degrees of freedom of the system, and dynamics, the movement of the system accounting for mass and inertia.

**Kinematics** Kinematics studies position, velocity, acceleration of human joint angles. The forward kinematics problem consists of finding the position of the end effector in Cartesian space in respect to the joint configurations of a skeletal system. This is a well-defined transformation, as a unique answer can be obtained in terms of the joint variables. However in movement control, the task is usually described in Cartesian space, which is different from the joint space in which the motor commands are applied and therefore, a coordinate transformation is required to find the solution in joint space. This problem, known as the *inverse kinematics* problem, is much more difficult than forward kinematics because the task space is usually the three dimensional world with six coordinates, while the number of joints is much greater. It is not easy to solve, but in a moment we will describe one of the methods for doing so.

**Dynamics** Given a kinematics solution, the next step is to calculate the forces and torques required to actually drive the skeletal system to move along the kinematics trajectory. Similar to kinematics, the dynamics step also involves two problems, the forward dynamics and the inverse dynamics. The forward problem is to determine the trajectories expressed in joint values, velocities and accelerations, given a series of forces and torques applied to the skeletal system. The inverse one is to determine the required time sequence of joint forces and torques to produce the desired time sequence of positions, velocities and accelerations. Just like the kinematics, its the inverse dynamics problem which is much harder than the forward dynamics problem, but this
is the one that needs to be solved to be able to make movements. Its hard because the equations are non-linear, and have to be solved using many iterations that take considerable time. This makes them an option for very fast silicon computers used by robots, but is a basic non-starter for humans.

<table>
<thead>
<tr>
<th>Equations</th>
<th>Robot</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinematics</td>
<td>easy</td>
<td>easy</td>
</tr>
<tr>
<td>Inverse kinematics</td>
<td>possible</td>
<td>hard</td>
</tr>
<tr>
<td>Dynamics</td>
<td>easy</td>
<td>easy</td>
</tr>
<tr>
<td>Inverse Dynamics</td>
<td>possible</td>
<td>hard</td>
</tr>
</tbody>
</table>

Table 6.2: Directly solving Newton’s equations poses problems for humans and robots. Robots can solve them in principle, but lots of computation is typically required.

**Surface Contact**  A special problem arises with *manipulative* motor control and that is the surface contact problem. One needs to control how the end effectors react when they make contact with the designated objects. In feedback control, the sensory feedback on the end effector takes about 150-200 ms to return to the brain, and another 150-200 ms for the motor commands to be sent back to the muscles. During this 300-400 ms, a rigid arm moving at a speed of 10 cm/sec would have penetrated into the object for 3-4 cm. Naturally the fingers would be damaged for most objects. Fortunately the human skeletal system is relatively light and has forms of compliance in the form of the passive spring-like properties of muscles and the cushioning of skin.

If the damage is to be avoided by conventional robot feedforward control models, such models need to be extremely accurate and work at very high bandwidth to send correct motor commands. But since noise and other disturbances are always present in the real world, motor commands are rarely able to bring the end effector to the accurate position. A potential way of avoiding damage is to use feedback control. But the time element means that the servo rate on any controller has to be very high. For example the servo rate for the SARCOS robotic systems hydraulically controlled arm is 10 KHz, a rate that is unrealistic for biological systems.
6.7 Human motor control

How does the brain solve these problems and generate motor commands to drive the limbs? The initial approaches used feedback. Adams first proposed the first closed-loop control system, which used sensory-motor feedback signals to minimize the errors between the desired output and actual output. However, the delays in pure feedback systems make it difficult for them to account for human performance. Thus subsequent proposals, derived from robotics research, calculated feed forward solutions based on inverse dynamics. More recent proposals emphasize internal models that include both feedback and feedforward control. The feedforward control directly obtains motor commands from an inverse model, given a motor task. In the feedback control, the actual sensory data is compared to the desired sensory data, and errors are used to improve the inverse model. However, these solutions are in the form of algorithms. To get further along, we have to introduce the special structure of the body. The dynamics equations are universal but different physical instantiations mean that there are different solutions possible. Very small insects can exploit the surface tension of water to navigate on its surface. Humans cannot do this but they have options that follow from the high strength to weight ratios of their limbs as well as their passive compliance and energy storage properties.

A central feature of the muscle systems of mammals in general and humans in particular is that they have a large number of physical structures that can store energy passively. Passive storage can also make the movement dynamics very energy efficient. Thus during the execution of the movement, the extra momentum when the end effector contacts the object is passively absorbed by the elastic muscles without hurting the limbs. For the cat it has been estimated that 70% of the energy used in walking comes from the natural elasticity in its muscles as the passive energy stored in the muscles

<table>
<thead>
<tr>
<th>Feature</th>
<th>Robot</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strength-to-wt ratio</td>
<td>100:1</td>
<td>1:1</td>
</tr>
<tr>
<td>Bandwidth-per-dof</td>
<td>10,000Hz</td>
<td>10Hz</td>
</tr>
<tr>
<td>Compliance</td>
<td>None</td>
<td>Extensive</td>
</tr>
</tbody>
</table>

Table 6.3: A comparison of basic features shows extensive differences between humans and robots.
can be used as an impetus for the next action step. A certain percentage of momentum energy is transformed into potential energy stored in muscles when the feet make contact with the ground. This potential energy is reused in the next cycle to push the feet forward.

**Equilibrium point control** The advantages of using stored energy was recognized early by Bernstein\(^5\) and led to the Feldman’s Equilibrium Point (EP) theoretical framework\(^6,7\) This states that limb movement can be controlled by shifting limb posture, represented as equilibrium, from one position to another. Subsequently, researchers in areas as physics, mechanical engineering, brain and cognitive science have successfully built passive walkers or walking models using damped springs as muscles. The first was Raibert’s robotic systems models\(^8,9\) consisting of a number of movement systems using damped springs. Most recently researchers at Cornell have demonstrated a bipedal walking system based on stored energy.\(^10\) This approach has also been recently studied in simulation by Gunther and Ruder\(^11\) who programmed a bipedal model consisting of 14 muscle-tendon complexes per leg and demonstrated stability margins in different planetary gravitational fields.

Although the above work has successfully modeled walking using the concept of passive energy, the methods are restricted to walking only, and not general enough to be applied to other human movements or to more complex situations. Gu proposed that humans can generate complex movements by first dividing them up into segments and then using EP models to calculate equilibrium points for each segment by simulating the movement. That is, the brain first plans the endpoint in musculo-skeletal space before the initiation of movements. She shows that this can be used not only to control simple movements such as reaching and walking, but also to use motor synergies to generate complex movements with composite goals, such as walking with reaching, sitting with balancing and other common abilities humans have.

The model has two main advantages. First, it separates motor planning from execution. During the planning stage, movements are divided into one or more segments. For each segment, the endpoint in joint space is calculated, given a task description in Cartesian space using the gradient descent method of.\(^12\) . Based on the simulations, the model is sufficiently general and robust to be applied to both simple and complicated movements. Second, she shows that EPs can be used in synergies to satisfy multiple
goals. If a motor control model represents a genuine human movement model, it should be easily extended to satisfy multiple goals, such as walking to pick up something, sitting down and reaching for something, as this is a common ability humans possess. She shows how our simulation method can be extended to multiple simultaneous movements.

The planning method also extends to complex movements such as walking. To do this one cycle of walking is divided into into four segments, and each segment has an end-point configuration. Natural and realistic walking is generated by shifting from one end-point to the next. Figure 6.12(a) gives the four equilibrium-points that are defined for one cycle of walking. For each leg, three damped springs are used to control the contraction and extension of three joints: hip, knee and ankle. Figure 6.12(b) presents four snapshots from the walking simulation. A comparison of the joint profiles obtained from simulation uses the human data from Boston Dynamics’ graphic human, shown in Fig. 6.12c, which plays back the walking data captured from human subjects. Figure 6.12d shows the data fitting results for hip, knee and ankle during one cycle of walking. Given that only four EPs are applied for the walking, claim the data fitting is quite promising.

Frames of reference and motor cortical organization  A skilled movement executed by a human is extremely complex. The six hundred muscles of the body have to be coordinated to carry out a sequence of segments. Think of an ice skater carrying out her program. Thus one has to expect that the cortical representation of movement would somehow reflect these complexities, which result in large numbers of muscles coordinated over time in precise contractions. The visual system is a useful comparison in this respect. Primate vision has very precise methods for quickly stabilizing gaze on targets in the visual field. Most of the connections use that stability to analyze the foveated area with elaborate precision. This point is brought home by Lennie, who showed that the connections in the neural hierarchy are heavily weighted towards the analysis of the pattern with respect to the fixated coordinate system. In an analogous way the movement patterns of the body have to be coded with respect to set frames of reference. In playing the piano, the finger movements heavily depend on the dimensions of the keys and the position of the seat to be in a very precise arrangement. Similarly in most sports, players use the position and orientation of the head as a coordinate system with which to access standard body movements. Golfers have to
Figure 6.12: Walking simulation, and comparison to human walking data. (a) Three abstract muscles are used to control the hip, knee and ankle. One cycle of walking is divided into four segments, and each segment has one equilibrium point. Walking is generated by shifting from one end-point to the next. (b) The snapshots from the walking simulation. (c) Boston Dynamics Inc. graphic human. The virtual agent walks by playing the walking data captured from human subjects. To quantitatively evaluate our model, we compared the joint profiles in simulation to the human data from the company.
6.7. HUMAN MOTOR CONTROL

keep their head still if they are going to compete on the PGA tour. Thus the motor system turns out to be unexpectedly similar to the visual system in that the emphasis is on elaborately coded patterns with respect to standard frames of reference. With this in mind, let us turn to the organization of the corticies that encode movements.

As shown in Fig. 6.13, the primary motor cortical area is located in a strip anterior to the central sulcus, a deep fold in the cortex. Like all of the cortex it represents a map of something, in this case neurons that can drive the body’s musculature. If stimulated independently, muscles in different parts of the body can be selectively activated. Thus one can draw a caricature of the anatomy on the motor strip representing the parts that responded. As shown by this depiction, some parts of the body are represented with increased resolution over others. The diagram is termed a homunculus, or ‘little man,’ for its human-like organization.

On the posterior side of the central sulcus is the all-important somatosensory area which also has a body map. Its job is to represent the sensed information as to the state of the joints et cetera. When this area is damaged, a person loses the sensation of the position of the damaged area, and has to rely on other senses, primarily vision, to maneuver it. These two areas, M1 and S1 respectively, form the ‘bottom’ of the cortex in an analogous way to

Figure 6.14: Graziano’s characterization of human hand-eye coordination movements. *(Light blue hand-to-mouth, Dark blue Reaching, Red Defense, Green Manipulation Purple Climbing)*

the role played by visual area V1.

As one might suspect after the example of the visual cortical hierarchy, the motor side is similarly organized into hierarchies representing abstractions needed to stitch together purposeful movement sequences. This organization has been studied by a number of researchers but notably Rizzolatti and colleagues. However our main focus is that of motor routines and its exposition centers around points that can be made at area M1, the lowest cortical level. The homunculus organization has stood the test of time as the basic finding that stimulation along the motor strip can produce movements predictably mapped to the body has been replicated many times. From the standpoint of what one might expect of movements though, it raises a number of issues, the chief one being that purposive movements usually coordinate many different muscles in the body and it is difficult to see how this is done is M1 is so distributed. For this reason some researchers have recently questioned the homunculus organization, notably the Graziano laboratory. By using larger than usual electrical stimulation in the monkey, they were able to produce multi-joint movements that could be classified in purposeful ways. Figure 6.14 shows his classification of the end points of stimulated movements. You can see immediately that these are larger, multi-joint movements that cluster in different regions.

If the cortex may be specially organized for purposive movements, they relevant neurons must pick a way to encode to the spinal cord. The main
issue to be tackled is the low cortical spike rates. These rates preclude using high bandwidth strategies employed by robots; there must be another way. One likely possibility is that the movement primitives that are needed are already in the spinal cord and there just needs to be a way of indexing them. These observation motivate the construction of a computational model of movement primitives.

6.8 Movement routines

Given that the spinal cord has primitives, the computational problem becomes finding a way that they can be characterized. Motivated by studies of more basic animal systems, a natural way of characterizing movements is in terms of temporal oscillations that are time-limited. This constraint in turn motivates the choice of a multi-scale Gabor dictionary as a means to decompose signals in both time and frequency space. Their advantages are 1) that human movements can be decomposed into time limited temporal segments owing to changing phasing and contact relationships, 2) human movements owing to inertia, tend to have a temporal smoothness that can be approximated nicely with sinusoids of varying frequencies, and 3) the reduced bandwidth required to drive the movement allows for abstract control strategies.

Let’s the terminology in Chen et al., defining a dictionary \( D = \phi_{\gamma}, \gamma \in \Gamma \) where \( \Gamma \) denotes the set of functions that we will use and \( \gamma \) is an index that ranges over the set. Then the signal \( x(t) \) can be encoded with the basis functions as,

\[
x(t) \approx \sum_{\gamma \in \Gamma} \alpha_{\gamma} \phi_{\gamma},
\]

Depending on the dictionary, the parameter \( \gamma \) can have the interpretation of indexing frequency (Fourier representation), of indexing time/scale jointly (Wavelet dictionaries), or of indexing time/frequency jointly (Gabor dictionaries).

Finding the optimal decomposition can be very computationally expensive, but fortunately for biological systems, the problem is made simpler by solving for an approximate decomposition. Generally, overcomplete dictionaries are used as they provide greater flexibility in capturing the data. Instead of a small set of general basis functions, there is a larger set of more specialized basis functions such that relatively few are required to represent
any particular signal. A sparse representation is a result of the presence of specialized basis functions that closely resemble the decomposed signal.

For the Gabor dictionary, \( \gamma = (\omega, \tau, \theta, \delta t) \), where \( \omega \in [0, \pi] \) represents the frequency, \( \tau \) represents the location in time, \( \theta \) is a phase, and \( \delta t \) is the duration. The atoms of the dictionary are defined as,

\[
\phi_{\gamma}(t) = e^{-\frac{(t - \tau)^2}{\delta t^2}} \cos(\omega(t - \tau) + \theta)
\]

For fixed \( \delta t \) and \( \theta \in \{0, \pi/2\} \), a discrete Gabor dictionary would be complete for \( \omega_k = k\Delta\omega \) and \( \tau_l = l\Delta\tau \), for sufficiently fine \( \Delta\omega \) and \( \Delta\tau \) (see\(^2\),\(^7\)), where \( k \) and \( l \) are integers. For our purposes, we would like the dictionary to be adapted by the learning algorithm so that it spans the space of movement signals in the data. To initialize the dictionary, we set \( \Delta\tau \) as a small multiple \((l)\) of the sampling rate of the signal, and \( \Delta\omega = 2 \), resulting in a sufficient resolution in the frequency space.

Figure 6.15: Gabor dictionary atoms across time. Scale, frequency, time shift, and phase are the four variable parameters. The learning algorithm controls all parameters except the phase which is set to 0 and \(-\pi/2\). Shown here are example atoms of the initial dictionary with time scaled from 0 to 1, where 0 is start of movement, and 1 indicates end of movement.

Figure 6.15 illustrates example dictionary atoms of the Gabor dictionary. The wavelets were initialized with dyadic scales, starting from a scale that
does not eliminate the smoothness of the Gabor (e.g. first column of figure 6.15), and ending in a scale that enables the Gabor to extend to the whole movement span (e.g. second column of figure 6.15). The exact structure of the dictionary at any time, however, would be determined by the learning algorithm, which is just a variant of the one used to find receptive fields for image data in Chapter three with the exception that the modification is carried out with respect to the parameters defining the curves rather than the curves themselves since this strategy is much more efficient. Although each dictionary atom is active only at a specific time, it is defined for the entire movement time (it is zero for the non-active part), to easily enable a scaled superposition of different atoms when reproducing a movement signal.

Figure 6.16: Reconstruction of the length of Gracilis muscle using Basis Pursuit for different number of coefficients. Bottom plot marks the particular dictionary atoms used in reconstructing the data. Each circle marks the time at which the atom was chosen and the ordinate value indicates the frequency of the atom. The color the circle indicates the relative magnitude of each coefficient, white being the highest.

Figure 6.16 illustrates the main result and that is that the length change of the *gracilis* muscle, one of the leg muscles used in walking, can be fit with extraordinary precision by using just 25 coefficients. With respect to
the model, these coefficients would be the $\alpha_\gamma$ in Equation 6.1. The functions denoted by $\phi_\gamma$ are assumed to be represented in the spinal cord. The reader will notice that this model has taken liberties with respect to the problem in that the length of the muscle was fit. What biology uses is not length but force. However since the muscles work by contracting and thus creating forces, there will be a relationship between force and length. Since it will be systematic and monotonic, fitting length is a demonstration that biology should be able to solve the slightly harder problem of determining forces by the same means. Figure 6.16 represents the traverse of just one muscle. Combining all the muscles' length changes allows the reconstruction of the movements of the many joints. Figure ?? shows one of them, the hip adduction. You can see that the movement is very close to the original data.

At this point, you must be wondering where the data came from that was fit with the Gabor basis function set. Human motion can be routinely captured using a motion capture suit that has key points that a surrounding array of cameras can identify and cooperate to recover their three dimensional coordinates. Next these points can be used to recover the movement of a model of the underlying skeleton, which has model muscles attached. These two stages are depicted in Fig. 6.18A.

Once the muscle temporal excursions have been fit, the result can be played back as shown in Fig. 6.18B. The fit is sufficiently exact that the replay cannot be easily distinguished from the original.
6.8. MOVEMENT ROUTINES

Figure 6.18: A) The lights attached to a body suit are recognized by cameras in the PhaseSpace system. These are then turned into muscle lengths using Stanford’s Open Simm software system. B) Human walking using muscle reconstruction with Gabor dictionary. The color represents the muscle length compared to its resting length, where red indicates that the length is greater, blue indicates that the length is less, and white indicates that the length is close to the resting length.
After all this development, you might be wondering how this formulation helps get around the basic non-linearity of the dynamics equations. The reasons is that the cost of the iterations can be amortized over the body’s lifetime. During the extended development period, one can explore possible movements and save the ones that work. Thus for any particular movement can be found using the cortex’s vast table look-up capacity. Small adjustments may need to be made but these are easy to handle.

6.9 Summary

The despite all the leverage gained by its parallelism, the brain’s neural circuitry is fundamentally slow, with the result that every possibility must be exploited to create a system that can perform in real-time. This chapter covered many important elements that allow this to happen.

1. The body that the brain is situated in can be regraded as a special-purpose computer that handles many very difficult aspects of the Newtonian dynamics,

2. Additionally the orientations of the body introduce critical frames-of-reference that make the subsequent computations easier,

3. These computations are further simplified by being purposive, allowing them to ignore aspects of the environment that are not relevant to immediate goals,

4. The spinal cord’s production of complex movements can be abstracted using a library of motor functions to make it accessible to the cortex.
Bibliography


