

Cracking the Neural Code

Information theory and efficient coding

Whereas the heart pumps blood and the lungs effect gas exchange, whereas the liver processes and stores chemicals and the kidney removes substances from the blood, the nervous system processes information.

Summary of Neurosciences Research Program work session, 1968

The goal of the 1968 Neurosciences Research Program meeting was to discuss how individual and groups of neurons process information. The meeting's summary, written by neuroscientists Theodore Bullock and Donald Perkel, does not push for any hard and fast conclusions. But it does lay out a wide world of possibilities for the representation, transformation, transmission and storage of information in the brain in a way that summarised the state of the field.

As the quote from their summary implies, ascribing the role of information processing to the brain seems as natural as saying the heart pumps blood. Even before 'information' became a part of everyday vocabulary in the twentieth century, scientists still spoke implicitly of the information that nerves convey, often in the language of 'messages' and 'signals'. An 1892 lecture to hospital employees, for example, explains that: 'There are fibres which convey messages from the various parts of the body to the brain' and that some of these fibres 'carry special kinds of messages as, for example, the nerves connected with the organs of special sense, which have been called the gateways of knowledge'. In the same vein, an 1870 publication describes the firing of motor neurons as 'a message of the will to the muscle' and even goes so far as to equate the nervous system with the dominant information-transmitting technology of the day: the telegraph.

But the investigation into how the nervous system represents information only started in earnest about 40 years before Bullock and Perkel's report, with the work of Edgar Adrian in the early twentieth century.

Adrian was in many ways the image of a prim and proper man of science. By the time he was born in London in 1889, his family had been in England for more than 300 years – a lineage that included a sixteenth-century surgeon and several reverends and members of government. As a student, his brilliance was regularly acknowledged by his teachers. In addition to his focus on medicine during his university studies, he displayed skill in art, particularly painting and drawing. As a lecturer at Cambridge, he worked long hours in the

lab and in the classroom. In his career as a physiologist, he was an undeniable success. At the age of 42 he won a Nobel Prize and in 1955 he was granted a title by Queen Elizabeth II, becoming Lord Adrian.

But behind these formal awards and accolades was a restless and chaotic man. Adrian was a thrill-seeker who liked climbing mountains and driving fast cars. He was happy to experiment on himself, including keeping a needle in his arm for two hours to try to measure muscle activity. He was known to play elaborate games of hide-and-seek with fellow students in the valleys of England's Lake District. As a professor he was equally elusive. He avoided unscheduled meetings by hiding in his lab, forcing any enquiring students to try to catch him on his bike ride home. He was temperamental and when he needed to think he'd perch himself on a shelf in a dark cabinet. His lab mates and his family described his movements as rapid, jerky and almost constant. His mind could be equally darting. Over the course of his career he studied many different questions in many different animals: vision, pain, touch and muscle control in frogs, cats, monkeys and more.

This inability to remain still, physically or mentally, may have been a key to his success. Through his varied studies on the activity of single nerves he was able to find certain general principles that would form the core of our understanding of the nervous system as a whole. In his 1928 book, *The Basis of Sensation*, Adrian explains his conclusions and the experiments that allowed him to reach them. The pages are peppered with talk of 'signals', 'messages' and even 'information', all mixed in with anatomical details of the nervous system and the technical challenges of capturing its activity. It was a mix of experimental advances and conceptual insights that would influence the field for decades to come.

In [Chapter 3](#), Adrian explains an experiment wherein he adds weight to a frog's muscle to see how the 'stretch' receptors that track the muscle's position would respond. Adrian recorded from the nerves that carry this signal from the receptors to the spinal cord. After applying different weights, Adrian summarised his findings as follows: 'The sensory message which travels to the central nervous system when a muscle is stretched ... consists of a succession of impulses of the familiar type. The frequency with which the impulses recur depends on the strength of the stimulus, but the size of each impulse does not vary.' This finding – that the size, shape or duration of an action potential emitted by these sensory neurons does not change, no matter how heavy or light the weight applied to the muscle is – Adrian referred to as the 'all-or-nothing' principle.

Examples of the 'all-or-nothing' nature of neural impulses reappear throughout the book. In different species, for different nerves carrying different messages, the story is always the same. Action potentials don't change based on the signal they're conveying, but their frequency can. The spikes of a neuron are thus like an army of ants – each as identical as possible, their power coming mainly from their numbers.

If the nature of an individual action potential is the same regardless of the strength or weakness of the sensory stimulus causing it, then one thing is certain: the size of the action potential does not carry information. With the contributions from Adrian, physiologists now felt comfortable embarking on a search for where exactly information was in nerves and how

it got transmitted.

There was only one problem: what *is* information? The blood that the heart pumps and the gases that lungs exchange are real, physical substances. They're observable, tangible and measurable. For as commonly as we use it, 'information' is actually a rather vague and elusive concept. A precise definition of the word does not easily come to mind for most people; it falls unfortunately into the 'know it when you see it' camp. Without a way to weigh information the way we can weigh fluids or gases, what hope could scientists have for a quantitative understanding of the brain's central purpose?

Between the time of Adrian's book and Perkel and Bullock's report, however, a quantitative definition of information had been found. It was born out of the scientific struggles of the Second World War and went on to transform the world in unexpected ways. Its application to the study of the brain was at times as rocky to execute as it was obvious to attempt.

* * *

Claude Shannon started at Bell Labs under a contract provided by the American military. It was 1941 and the National Defense Research Committee wanted scientists working on wartime technology. The seriousness of the work didn't dampen Shannon's naturally playful tendencies, though. He enjoyed juggling and while at Bell Labs was known to juggle around campus while riding a unicycle.

Born in a small town in the American Midwest, Shannon grew up following his curiosity about all things science, mathematics and engineering anywhere it took him. As a child he played with radio parts and enjoyed number puzzles. As an adult he created a mathematical theory of juggling and a flame-powered Frisbee. He enjoyed chess and building machines that could play chess. A constant tinkerer, he made many gadgets, some more productive than others. On his desk at Bell Labs, for example, he kept an 'Ultimate Machine': a box with a switch that, when flipped on, caused a mechanical hand to reach out and flip it back off.¹

For his master's degree, Shannon wrote a 72-page thesis entitled 'A symbolic analysis of relay and switching circuits' that would revolutionise electrical engineering. For his PhD, he turned his mathematical eye toward biology, working on 'An Algebra for Theoretical Genetics'. But his topic at Bell Labs was cryptography. How to safely encode messages that would be transmitted through land, air and water was a natural topic of concern for the military. Bell Labs was a hub of cryptography research and even hosted the renowned code-cracker Alan Turing during Shannon's time there.

All this work on codes and messages got Shannon thinking broadly about the concept of communication. During the war, he proposed a method for understanding message-sending mathematically. Because of the necessary secrecy around cryptography research, however, his ideas were kept classified. In 1948, Shannon was finally able to publish the work and 'A mathematical theory of communication' became the founding document of a new field: information theory.

Shannon's paper describes a very generic communication system consisting of five simple

parts. The first is an information *source*, which produces the message that will be sent. The next is the *transmitter*, which is responsible for encoding the message into a form that can be sent across the third component, the *channel*. On the other end of the channel, a *receiver* decodes the information back into its original form and it is sent to its final *destination* (Figure 16).

In this framework, the medium of the message is irrelevant. It could be songs over radio waves, words on a telegraph or images through the internet. As Shannon says, the components of his information-sending model are ‘suitably idealised from their physical counterparts’. This is possible because, in all of these cases, the fundamental problem of communication remains the same. It is the problem of ‘reproducing at one point either exactly or approximately a message selected at another point’.

Shannon’s Communication System

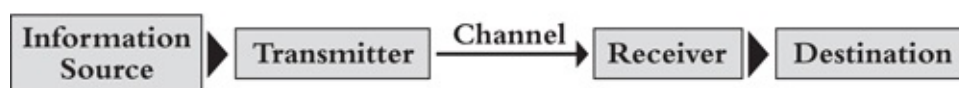


Figure 16

With this simple communication system in mind, Shannon aimed to formalise the study of information transmission. But, to mathematically approach the question of how information is communicated, he first had to define information mathematically. Building on previous work, Shannon discusses what desirable properties a measure of information should have. Some are practical: information shouldn’t be negative, for example, and its definition should be easy to work with mathematically. But the real constraint came from the need to capture an intuition about information: its reliance on a code.

Imagine a school where all students wear uniforms. Seeing a student show up in the same outfit every day provides very little information about their mood, their personality or the weather. On the other hand, in a school without uniforms, clothing choice has the ability to convey all this information and more. To someone wondering about the current temperature, for example, seeing a student in a sundress instead of a sweater can go a long way towards relieving that curiosity. In this way, clothing can be used as a code – it is a transmittable set of symbols that conveys meaning.

The reason the uniformed students can’t carry this information is that a code requires options. There need to be multiple symbols in a code’s vocabulary (in this case, multiple outfits in a student’s wardrobe), each with their own meaning, for any of the symbols to have meaning.

But it’s not just the number of symbols in a code that matters – it’s also how they’re used. Let’s say a student has two outfits: jeans and a T-shirt or a suit. If the student wears jeans and a T-shirt 99 per cent of the time, then there is not much information that can be gained from this wardrobe choice. You wouldn’t even need to see the student to be almost certain of what they’re wearing – it’s essentially a uniform. But the one day in a hundred where they show up in a suit tells you something important. It lets you know that day is somehow special. What

this shows is that the rarer a symbol's use, the more information it contains. Common symbols, on the other hand, can't communicate much.

Shannon wanted to capture this relationship between a symbol's use and its information content. He therefore defined a symbol's information content in terms of the probability of it appearing. Specifically – to make the amount of information decrease as the probability of the symbol increases – he made a symbol's information depend on the *inverse* of its probability. Because the inverse of a number is simply one divided by that number, a higher probability means a lower 'inverse probability'. In this way, the more frequently the symbol is used, the lower its information will be. Finally, to meet his other mathematical constraints, he took the logarithm of this value.

A logarithm, or 'log', is defined by its *base*. To take the base-10 log of a number, for example, you would ask: 'To what power must I raise 10 in order to get this number?' The base-10 log of 100 (written as $\log_{10}100$), is therefore 2, because 10 to the power of 2 (*i.e.*, 10×10) is 100. The base-10 log of 1,000 is thus 3. And the base-10 log of something in between 100 and 1,000 is in between 2 and 3.

Shannon decided to use a base of *two* for his definition of information. To calculate the information in a symbol you must therefore ask: 'To what power must I raise two in order to get the inverse of the symbol's probability?' Taking our student's outfit of jeans and T-shirt as a symbol that appears with 0.99 probability, its information content is $\log_2(1/0.99)$, which comes out to about 0.014. The suit that appears with only 0.01 probability, on the other hand, has an information content of $\log_2(1/0.01)$ or roughly 6.64. Again, the lower the probability, the higher the information.²

But Shannon was interested in more than just the information in a single symbol – he wanted to study the information content of a code. A code is defined by its set of symbols and how frequently each is used. Shannon therefore defined the total information in a code as the sum of the information of all its symbols. Importantly, this sum is weighted – meaning the information from each symbol is multiplied by how frequently that symbol is used.

Under this definition, the student's clothing code would have a total information content of 0.99×0.014 (from the jeans and T-shirt) + 0.01×6.64 (from the suit) = 0.081. This can be thought of as the average amount of information we would receive each day by seeing the student's outfit. If the student chose instead to wear their jeans 80 per cent of the time and their suit the other 20 per cent, their code would be different. And the average information content would be higher: $0.80 \times \log_2(1/0.80) + 0.20 \times \log_2(1/0.20) = 0.72$.

Shannon gave the average information rate of a code a name. He called it entropy. The official reason he gives for this is that his definition of information is related to the concept of entropy in physics, where it serves as a measure of disorder. On the other hand, Shannon was also known to claim – perhaps jokingly – that he was advised to call his new measure entropy because 'no one understands entropy' and therefore Shannon would likely always win arguments about his theory.

Shannon's entropy captures a fundamental trade-off inherent in maximising information. Rare things carry the most information, so you want as much of them as possible in your

code. But the more you use a rare symbol, the less rare it becomes. This fight fully defines the equation for entropy: decreasing the probability of the symbol makes the log of its inverse go up – a positive contribution to the information. But this number is then multiplied by that very same probability: this means that decreasing a symbol's probability makes its contribution to information go down. To maximise entropy, then, we must make rare symbols as common as possible but no commoner.

Shannon's use of a base-two log makes the unit of information the *bit*. Bit is short for binary digit and, while Shannon's paper sees the first known use of the word, he did not coin it (Shannon credits his Bell Labs colleague John Tukey with that honour).³ The bit as a unit of information has a helpful and intuitive interpretation. Specifically, the average number of bits in a symbol is equal to the number of yes-or-no questions you need to ask in order to get that amount of information.

For example, consider trying to find out the season in which someone was born. You may start by asking: 'Is it a transitional season?' If they say yes, you may then ask: 'Is it spring?' If they say yes to that, you have your answer; if they say no, you still have your answer: autumn. If they said no to the first question you'd follow the opposite route – asking if they were born in summer, *etc.* No matter the answer, it takes two yes-or-no questions to get it. Shannon's entropy equation agrees. Assuming people are equally likely to be born in each season, then each of these season 'symbols' will be used 25 per cent of the time. The information in each symbol is thus $\log_2(1/0.25)$. This makes the average bits per symbol equal to two – the same as the number of questions.

Part of designing a good communication system is designing a code that packs in a lot of information per symbol. To maximise the average information that a symbol in a code provides, we need to maximise the code's entropy. But, as we saw, the definition of entropy has an inherent tension. To maximise it, rare symbols need to be the norm. What is the best way to satisfy this seemingly paradoxical demand? This tricky question turns out to have a simple answer. To maximise a code's entropy, each of its symbols should be used the exact same amount. Have five symbols? Use each one-fifth of the time. A hundred symbols? Each should have a probability of 1/100th. Making each and every symbol equally likely balances the trade-off between rare and common communication.

What's more, the more symbols a code has the better. A code with two symbols, each used half the time, has an entropy of one bit per symbol (this makes sense according to our intuitive definition of a bit: thinking of one symbol as standing in for 'yes' and the other for 'no', each symbol answers one yes-or-no question). On the other hand, a code with 64 symbols – each used equally – has an entropy of six bits per symbol.

As important as a good code is, encoding is only the start of a message's journey. According to Shannon's conception of communication, after information is encoded it still needs to be sent through a channel on the way to its destination. Here is where the abstract aims of message-sending meet the physical limits of matter and materials.

Consider the telegraph. A telegraph sends messages via short pulses of electric current passed over wires. The patterns of pulses – combinations of shorter 'dots' and longer

‘dashes’ – define the alphabet. In American Morse code, for example, a dot followed by a dash indicates the letter ‘A’, two dots and a dash means ‘U’. The physical constraints and imperfections of the wires carrying these messages, especially those sent long distances or under oceans, put a limit on the speed of information. Telegraph operators who typed too quickly were at risk of running their dots and dashes together, creating an unintelligible ‘hog Morse’ that would be worthless to the receiver of it. In practice, operators could safely send around 100 letters per minute on average.

To create a practical measure of information rate, Shannon combined the inherent information rate of a code with the physical transmission rate of the channel it is sent over. For example, a code that provides five bits of information per symbol and is sent over a channel that can send 10 symbols per minute would have a total information rate of 50 bits per minute. The maximum rate at which information can be sent over a channel without errors is known as the channel’s capacity.

Shannon’s publication enforced a clear structure on a notoriously nebulous concept. In this way, it set the stage for the increasing objectification of information in the decades to come. The immediate effects of Shannon’s work on information processing in the real world, however, were slight. It took more than two decades for the technology that makes information transmission, storage and processing a constant part of everyday life to come about. And it took engineers time to figure out how to harness Shannon’s theory for practical effect in these devices. Information theory’s impact on biology, however, came much quicker.

* * *

The first application of information theory to biology was itself a product of war. Henry Quastler, an Austrian physician, was living and working in the US during the time of the Second World War. His response to the development of the atomic bomb was one of horror – and action. He left his private practice to start doing research on the medical and genetic effects of nuclear bombs. But he needed a way to quantify just how the information encoded in an organism was changed by exposure to radiation. ‘A godsend, these formulas, splendid! I can go on now,’ Quastler is said to have remarked upon learning about Shannon’s theory. He wrote a paper in 1949 – just a year after Shannon’s work was published – entitled ‘The information content and error rate of living things’. It set off the study of information in biology.

Neuroscience was not slow to follow. In 1952, Warren McCulloch and physicist Donald MacKay published ‘The limiting information capacity of a neuronal link’. In this paper, they derive what they consider to be the most optimistic estimate of how much information a single neuron could carry. Based on the average time it takes to fire an action potential, the minimum time needed in between firing and other physiological factors, MacKay and McCulloch calculated an upper bound of 2,900 bits per second.

MacKay and McCulloch were quick to stress that this doesn’t mean neurons actually *are* conveying that much information, only that under the best possible circumstances they could. After their paper, many more publications followed, each aiming to work out the true coding

capacity of the brain. So awash in attempts was the field that in 1967 neuroscientist Richard Stein wrote a paper both acknowledging the appeal of information theory for quantifying nervous transmission but also lamenting the ‘enormous discrepancies’ that have resulted from its application. Indeed, in the work that followed MacKay and McCulloch’s, estimates ranged from higher than their value – 4,000 bits per second per neuron – to significantly lower, as meagre as one-third of a bit per second.

This diversity came, in part, from different beliefs about how the parts and patterns of nervous activity should be mapped on to the formal components of Shannon’s information theory. The biggest question centred on how to define a symbol. Which aspects of neural activity are actually carrying information and which are incidental? What, essentially, is the neural code?

Adrian’s original finding – that it is not the height of the spike that matters – still held strong.⁴ But even under this constraint, options abounded. Starting from the basic unit of an action potential, scientists were still able to devise many conceivable codes. MacKay and McCulloch began by thinking of the neural code as composed of only two symbols: spike or no spike. At each point in time a neuron would send one or the other symbol. But after calculating the information rate of such a code, MacKay and McCulloch realised they could do better. Thinking instead of the *time between spikes* as the code allowed a neuron to transmit much more information. In this coding scheme, if there were a 20-millisecond gap in between two spikes, this would symbolise something different from a 10-millisecond gap. It’s a scheme that creates many more possible symbols and it was with this style of coding that they made their estimate of 2,900 bits per second.

Stein, in his attempt to clean up the cacophony of codes on offer at the time, focused on a third option for neural coding – the one that came from Adrian himself. Adrian, after establishing that action potentials don’t change as the stimulus does, claimed that: ‘In fact, the only way in which the message can be made to vary at all is by a variation in the total number of the impulses and in the frequency with which they recur.’ This style of coding – where it is the number of spikes produced in a certain amount of time that serves as the symbol – is known as frequency or rate-based coding. In his 1967 paper, Stein argues for the existence of a rate-based code and highlights its benefits, including higher error tolerance.

But the debate over what the true neural code is did not end with Stein in 1967. Nor did it end with Bullock and Perkel’s meeting on information coding in the brain a year later. In fact, in their report on that meeting, Bullock and Perkel include an Appendix that lays out dozens of possible neural codes and how they could be implemented.

In truth, neuroscientists continue to spar and struggle over the neural code to this day. They host conferences centred on ‘Cracking the neural code’. They write papers with titles like ‘Seeking the neural code’, ‘Time for a new neural code?’ and even ‘Is there a neural code?’ They continue to find good evidence for Adrian’s original rate-based coding, but also some against it. Identifying the neural code can seem a more distant goal now than when MacKay and McCulloch wrote their first musings on it.

In general, some evidence of rate-based coding can be found in most areas of the brain. The

neurons that send information from the eye change their firing frequency based on the intensity of light. Neurons that encode smell fire in proportion to the concentration of their preferred odour. And, as Adrian showed, receptors in the muscle and those in the skin fire more with more pressure applied to them. But some of the strongest evidence for other coding schemes comes from sensory problems that require very specific solutions.

When localising the source of a sound, for example, precise timing matters. Because of the distance between the two ears, sound coming from the left or right side will hit one ear just before it hits the other. This gap between the times of arrival at each ear – sometimes as brief as only a few millionths of a second – provides a clue for calculating where the sound came from. The medial superior olive (MSO), a tiny cluster of cells located right in between the two ears, is responsible for performing this calculation.

The neural circuit that can carry this out was posited by psychologist Lloyd Jeffress in 1948 and has been supported by many experiments since. Jeffress' model starts with information coming from each ear in the form of a temporal code – that is, the exact timing of the spikes matters. In the MSO, cells that receive inputs from each ear compare the relative timing of these two inputs. For example, one cell may be set up to detect sounds that arrive at both ears simultaneously. To do so, the signals from each ear would need to take the exact same amount of time to reach this MSO cell. This cell then fires when it receives two inputs at the exact same time and this response indicates that the sound hit both ears at the same time (see Figure 17).

The cell next to this one, however, receives slightly asymmetric inputs. That is, the nerve fibre from one ear needs to travel a *little* farther to reach this cell than the nerve from the other ear. Because of this, one of the temporal signals gets delayed. The extra length the signal travels determines just how much extra time it takes. Let's say the signal from the left ear takes an extra 100 microseconds to reach this MSO cell. Then, the only way this cell will receive two inputs at once is if the sound hits the left ear 100 microseconds before it hits the right. Therefore, this cell's response (which, like the other cell, only comes when it receives two inputs at once) would signal a 100-microsecond difference.

Continuing this pattern, the next cell may respond to a 200-microsecond difference, the one after that 300 microseconds and so on. In total, the cells in the MSO form a map wherein those that signal short arrival time differences fall at one end and those that signal long differences fall at the other. In this way, a temporal code has been transformed into a *spatial* code: the position of the active neuron in this map carries information about the source of the sound.

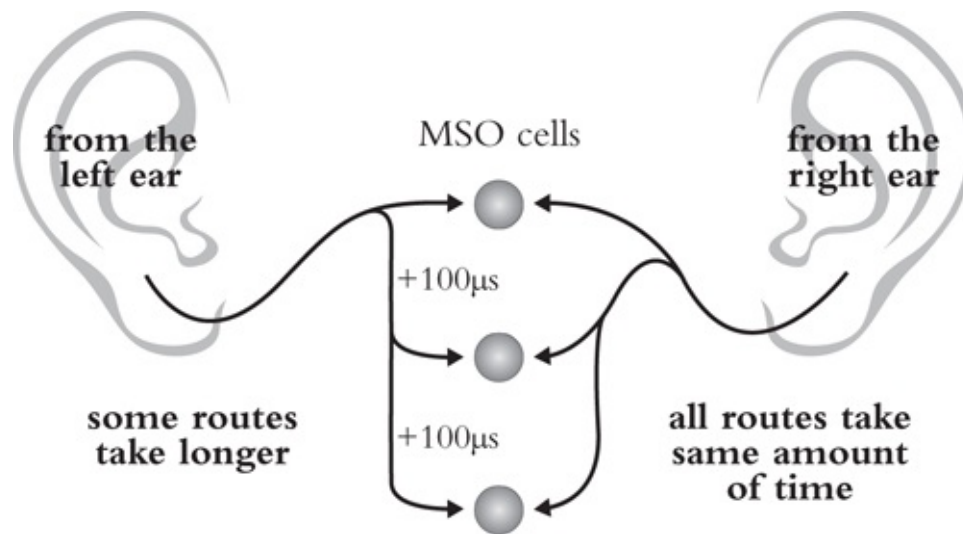


Figure 17

For the question of why the neural code is such an enigma, the most likely answer – as with so many questions of the brain – is because it's complicated. Some neurons, in some areas of the brain, under some circumstances, may be using a rate-based code. Other neurons, in other times and places, may be using a code based on the timing of spikes, or the time in between spikes, or some other code altogether. As a result, the thirst to crack *the* neural code will likely never be quenched. The brain, it seems, speaks in too many different languages.

* * *

Evolution did not furnish the nervous system with one single neural code, nor did it make it easy for scientists to find the multitude of symbols it uses. But, according to British neuroscientist Horace Barlow, evolution did thankfully provide one strong guiding light for our understanding of the brain's coding scheme. Barlow is known as one of the founders of the efficient coding hypothesis, the idea that – no matter what code the brain uses – it is always encoding information *efficiently*.

Barlow was a trainee of Lord Adrian. He worked with him – when he could find him – as a student at Cambridge in 1947. Barlow had always had a keen interest in physics and mathematics but, to be practical, chose to study medicine.⁵ Yet throughout his studies he recognised how the influence from more quantitative subjects could drive questions in biology. It was a trait he considered in contrast to his mentor: '[Adrian] was not at all theoretically based; his attitude was that we had the means of recording from nerve fibres and we should just see what happens.'

Quickly taken in by Shannon's equations when they came about, Barlow made several early contributions to the study of information in the brain. Rather than simply counting bits per second, however, Barlow's use of information theory went deeper. The laws of information, in some respects, are as fundamental and constraining of biology as the laws of physics. From Barlow's perspective, these equations could thus do more than merely *describe* the brain as it is, but rather *explain* how it came to be. So certain of its importance to neuroscience, Barlow compared trying to study the brain without focusing on information

processing to trying to understand a wing without knowing that birds fly.

Barlow came to his efficient coding hypothesis by combining reflections on information theory with observations of biology. If the brain evolved within the constraints of information theory – and evolution tends to find pretty good solutions – then it makes sense to conclude that the brain is quite good at encoding information. ‘The safe course here is to assume that the nervous system is efficient,’ Barlow wrote in a 1961 paper. If this is true, any puzzle about why neurons are responding the way they are may be solved by assuming they are acting efficiently.

But what does efficient information coding look like? For that, Barlow focused on the notion of redundancy. In Shannon’s framework, ‘redundancy’ refers to the size of the gap between the highest possible entropy a given set of symbols could have and the entropy they actually have. For example, if a code has two symbols and uses one of them 90 per cent of the time and the other only 10 per cent, its entropy is not as high as it could be. Sending the same symbol nine out of ten times is redundant. As we saw earlier, the code with the highest entropy would use each of those symbols 50 per cent of the time and would have a redundancy of zero. Barlow believed efficient brains reduce their redundancy as much as possible.

The reason for this is that redundancy is a waste of resources. The English language, as it turns out, is incredibly redundant. A prime example of this is the letter ‘q’, which is almost always followed by ‘u’. The ‘u’ adds little if any information once we see the ‘q’ and is therefore redundant. The redundancy of English means we could, in theory, be conveying the same amount of information with far fewer letters. In fact, in his original 1948 paper, Shannon estimated the redundancy of written English to be about 50 per cent. This is why, for example, ppl cn stll rd sntncls tht hv ll th vwls rmvd.⁶

In the nervous system, redundancy can come in the form of multiple neurons saying the same thing. Imagine one neuron represents the letter ‘q’ and another the letter ‘u’. The sight of ‘qu’ would thus make both these neurons fire. But if these two letters frequently appear together in the world, it would be more efficient of the brain to use just a single neuron to respond to them.

Why should it matter if the brain is efficient with its encoding? One reason is energy costs. Every time a neuron fires a spike, the balance of charged particles inside and outside the cell gets thrown off. Restoring this balance takes energy: little pumps in the cell membrane have to chuck sodium ions out of the cell and pull potassium ones back in. Building up neurotransmitters and expelling them from the cell with each spike also incurs a cost. In total, it’s estimated that up to three-fourths of the brain’s energy budget goes towards sending and receiving signals. And the brain – using 20 per cent of the body’s energy while accounting for only 2 per cent of its weight – is the most energetically expensive organ to run. With such a high energy bill, it makes sense for the brain to be economical in how it uses its spikes.

But to know how to send information efficiently, the brain needs to know what kind of information it normally needs to send. In particular, the brain needs to somehow determine when the information it is receiving from the world is redundant. Then it could simply not

bother sending it on. This would keep the neural code efficient. Does the nervous system have the ability to track the statistics of the information it's receiving and match its coding scheme to the world around it? One of Lord Adrian's own findings – adaptation – suggests it does.

In his experiments on muscle stretch receptors, Adrian noticed that 'there is a gradual decline in the frequency of the discharge under a constant stimulus'. Specifically, while keeping the weight applied to the muscle constant, the firing rate of the nerve would decrease by about half over 10 seconds. Adrian called this phenomenon 'adaptation' and defined it as 'a decline in excitability caused by the stimulus'. Noticing the effect in several of his experiments, he devoted a whole chapter to the topic in his 1928 book.

Adaptation has since been found all over the nervous system. For example, the 'waterfall effect' is a visual illusion wherein the sight of movement in one direction then causes stationary objects to appear as though they are moving in the opposite direction. It's so-named because it can happen after staring at the downward motion of a waterfall. The effect is believed to be the result of adaptation in the cells that represent the original motion direction: with these cells silenced by adaptation, our perception is biased by the firing of cells that represent the opposite direction.

In his 1972 paper, Barlow argues for adaptation as a means of increasing efficiency: 'If sensory messages are to be given a prominence proportional to their informational value, mechanisms must exist for reducing the magnitude of representation of patterns which are constantly present, and this is presumably the underlying rationale for adaptive effects.'

In other words – specifically, in the words of information theory – if the same symbol is being sent across the channel over and over, its presence no longer carries information. Therefore, it makes sense to stop sending it. And that is what neurons do: they stop sending spikes when they see the same stimulus over and over.

Since the time that Barlow made the claim that cells should adapt their responses to the signals they're receiving, techniques for tracking how neurons encode information have developed that allow for more direct and nuanced tests of this hypothesis. In 2001, for example, computational neuroscientist Adrienne Fairhall, along with colleagues at the NEC Research Institute in Princeton, New Jersey, investigated the adaptive abilities of visual neurons in flies.

For their experiment, the researchers showed the flies a bar moving left and right on a screen. At first, the bar's motion was erratic. At one moment it could be moving very quickly leftwards, and at the next it could go equally fast towards the right, or it could stay in that direction, or it could slow down entirely. In total, its range of possible speeds was large. After several seconds of such chaos, the bar then calmed down. Its movement became more constrained, never going too quickly in either direction. Over the course of the experiment, the bar switched between such periods of erratic and calm movement several times.

Looking at the activity of the neurons that respond to motion, the researchers found that the visual system rapidly adapts its code to the motion information it's currently getting. Specifically, to be an efficient encoder, a neuron should always fire at its peak firing rate for

the fastest motion it sees and at its lowest for the slowest.⁷ Thinking of different rates of firing as different symbols in the neural code, spreading the firing rates out this way ensures that all these symbols get used roughly equally. This maximises the entropy of the code.

The problem is that the fastest motion during the period of calm is much slower than the fastest motion during the more erratic period. This means that the same exact speed needs to map to two different firing rates depending on the context it appears in. Strange as it is, this is just what Fairhall and colleagues saw. During the calm period, when the bar was moving at its fastest speed, the neuron fired at over 100 spikes per second. Yet when that same speed occurred during the erratic period, the neuron only fired about 60 times per second. To get the neuron back up to 100 spikes per second during the erratic period, the bar needed to move 10 times as fast.

In addition, the researchers were able to quantify the amount of information carried by a spike before and after the switch between these two types of motion. During the erratic period, the information rate was around 1.5 bits per spike. Immediately after the switch to calm movement, it dropped to just 0.8 bits per spike: the neuron, having not yet adapted to the new set of motion it was seeing, was an inefficient encoder. After just a fraction of a second of exposure to the calmer motion, however, the bits per spike went right back up to 1.5. The neuron needed just a small amount of time to monitor the range of speeds it was seeing and adapt its firing patterns accordingly. This experiment shows that, just as Barlow's efficient coding theory suggests, adaptation ensures that all types of information are encoded efficiently.

Neuroscientists also believe that the brain is built to produce efficient encodings on much longer timescales than the seconds to minutes of a sensory experience. Through both evolution and development, an organism has a chance to sample its environment and adapt its neural code to what is most important to it. By assuming that a given area of the brain is best-suited to represent relevant information as efficiently as possible, scientists are attempting to reverse engineer the evolutionary process.

The 30,000 nerves that leave the human ear, for example, respond to different types of sounds. Some of the neurons prefer short blips of high-pitched noises, others prefer low-pitched noises. Some respond best when a soft sound gets louder, others when a loud sound gets softer and others still when a soft sound gets louder then softer again. Overall, each nerve fibre has a complex pattern of pitches and volumes that best drive its firing.

Scientists know, for the most part, *how* the fibres end up with these responses. Tiny hairs connected to cells in the inner ear move in response to sounds. Each cell responds to a different pitch based on where it is in a small, spiral-shaped membrane. The nerve fibres that leave the ear get inputs from these hairy cells. Each fibre combines pitches in its own way to make its unique, combined response profile.

What is less clear, however, is *why* the fibres have these responses. That's where ideas from information theory can help.

If the brain is indeed reducing redundancy as Barlow suggests, then only a small number of neurons should be active at a time. Neuroscientists refer to this kind of activity as 'sparse'.⁸

In 2002, computational neuroscientist Michael Lewicki asked whether the response properties of auditory nerves could be the result of the brain enforcing a sparse code – one specifically designed for the sounds an animal needs to process.

To answer this, he first had to gather a collection of different natural sounds. One set of sounds came from a CD of vocalisations made by rainforest animals such as bats, manatees and marmosets; another was a compilation of ‘background’ noises like crunching leaves and snapping twigs; and the third was from a database of human voices reading English sentences.

Lewicki then used an algorithm to decompose these complex sounds into a dictionary of short sound patterns. The goal of the algorithm was to find the best decomposition – that is, one that can recreate each full, natural sound using as few of the short sound patterns as possible. In this way, the algorithm was seeking a sparse code. If the brain’s auditory system evolved to sparsely encode natural sounds, the sound patterns preferred by the auditory nerves should match those found by the algorithm.

Lewicki found that creating a dictionary from just the animal noises alone produced sound patterns that didn’t match the biology. Specifically, the patterns the algorithm produced were too simple – representing just pure tones rather than the complex mix of pitches and volumes that human and animal auditory nerves tend to prefer. Applying the algorithm to a mix of animal noises and background sounds, however, did mimic the biology. This suggests that the coding scheme of the auditory system is indeed matched to these environmental sounds, allowing it to encode them efficiently. What’s more, Lewicki found that a dictionary made from human speech also reproduced the sound profiles preferred by biology. Lewicki took this as evidence for the theory that human speech evolved to make best use of the existing encoding scheme of the auditory system.⁹

* * *

In 1959, Barlow presented his ideas about the brain’s information-processing properties to a group of sensory researchers gathered at MIT. When the proceedings of this meeting were translated into Russian for a Soviet audience, Barlow’s contribution was conspicuously cut out. The Soviets, it turned out, had a problem with the use of information theory to understand the brain. Considered part of the ‘bourgeois pseudoscience’ of cybernetics, it ran counter to official Soviet philosophy by equating man with machine. Soviet leaders – and the sometimes scared scientists under their rule – openly critiqued this attitude as a foolish product of American capitalism.

Though unique in its political motivations, the Soviets’ denunciation was far from the only critique of information theory in biology. In 1956, a short article entitled ‘The bandwagon’ cautioned against the over-excited application of information theory in fields such as psychology, linguistics, economics and biology. ‘Seldom do more than a few of nature’s secrets give way at one time. It will be all too easy for our somewhat artificial prosperity to collapse overnight when it is realised that the use of a few exciting words like *information*, *entropy*, *redundancy*, do not solve all our problems.’ The article was written by Shannon

himself, just eight years after he unleashed information theory on to the world.

Concerns about how apt the analogy is between Shannon's framework and the brain have even come from the very scientists doing the analogising. In a 2000 article, Barlow warned that 'the brain uses information in different ways from those common in communication engineering'. And Perkel and Bullock, in their original report, made a point of not committing themselves fully to Shannon's definition of information but, rather, treating the concept of 'coding' in the brain as a metaphor that may have varying degrees of usefulness.

The caution is warranted. A particularly tricky part of Shannon's system to map to the brain is the decoder. In a simple communication system, the receiver gets the encoded message through the channel and simply reverses the process of encoding in order to decode it. The recipient of a telegraph message, for example, would use the same reference table as the sender to know how to map dots and dashes back to letters. The system in the brain, however, is unlikely to be so symmetric. This is because the only 'decoders' in the brain are other neurons, and what they do with the signal they receive can be anyone's guess.

Take, for example, encoding in the retina. When a photon of light is detected, some of the cells in the retina (the 'on' cells) encode this through an increase in their firing rate while another set of cells (the 'off' cells) encodes it by decreasing their firing. If this joint up-down change in firing is the symbol the retina has designated to indicate the arrival of a photon, we may assume this is also the symbol that is 'decoded' by later brain areas. However, this does not seem to be the case.

In 2019, a team of researchers from Finland genetically modified the cells in a mouse's retina. Specifically, they made the 'on' cells less sensitive to photons. Now, when a photon hit, the 'off' cells would still decrease their firing, but the 'on' cells may or may not increase theirs. The question was: which set of cells would the brain listen to? The information about the photon would be there for the taking if the 'off' cells were decoded. Yet the animals didn't seem to use it. By assessing the animal's ability to detect faint lights, it appeared that the brain was reading out the activity of the 'on' cells alone. If those cells didn't signal that a photon was detected, the animal didn't respond. The scientists took this to mean that the brain is not, at least in this case, decoding all of the encoded information. It ignores the signals the 'off' cells are sending. Therefore, the authors wrote, 'at the sensitivity limit of vision, the decoding principles of the brain do not produce the optimal solution predicted by information theory'. Just because scientists can spot a signal in the spikes doesn't mean it has meaning to the brain.

There are many reasons this might be. One important one is that the brain is an *information-processing* machine. That is, it does not aim to merely reproduce messages sent along it, but rather to transform them into action for the animal. It is performing computations on information, not just communicating it. Expectations about how the brain works based solely on Shannon's communication system therefore miss this crucial purpose. The finding that the brain is not optimally transmitting information does not necessarily indicate a flaw in its design. It was just designed for something else.

Information theory, invented as a language for engineered communication systems,

couldn't be expected to translate perfectly to the nervous system. The brain is not a mere telephone line. Yet parts of the brain do engage in this more basic task of communication. Nerves do send signals. And they do so through some kind of code based on spike rates or spike times or spike something. To glance at the brain from the vantage point of information theory, then, is a sensible endeavour – one that has yielded many insights and ideas. Stare too long, though, and the cracks in the analogy become visible. This is the reason for wariness. As a metaphor, the relationship between a communication system and the brain is thus most fruitful when not overextended.

Notes

- 1 Marvin Minsky, one of the authors of the *Perceptrons* book from Chapter 3, was working under Shannon at the time and is credited with the design of the Ultimate Machine. Shannon reportedly convinced Bell Labs to produce several of them as gifts for AT&T executives.
- 2 More on probability and its history in Chapter 10.
- 3 The base-two log wasn't the only option for information. Prior to Shannon's work, his colleague Ralph Hartley posited a definition of information using a base-10 log, which would've put information in terms of 'decimal digits or 'dits' instead of bits.
- 4 Full disclosure: some modern neuroscientists are exploring the idea that action potentials actually *do* change in certain ways depending on what inputs the cell gets and that these changes *could* be part of the neural code. Science is never set in stone.
- 5 Barlow credits his mother Nora, granddaughter of Charles Darwin, with his interest in science.
- 6 In case you couldn't: 'people can still read sentences that have all the vowels removed'. Texting and tweeting are also great ways to see just how many letters can be removed from a word before causing problems.
- 7 Technically if the neuron has a preferred direction of motion – that is, it fires most strongly for, say, rightwards motion – it should fire at its peak rate for high speed in that direction and its lowest rate for high speed in the opposite direction. But the principle remains the same regardless.
- 8 Among neuroscientists, the 'grandmother cell' is considered the mascot of sparse coding. This fictional neuron is meant to be the one and only cell to fire when you see your grandmother (and fires in response to nothing else). Such an extreme example of efficient coding was devised by Jerome Lettvin (the frog guy from the last chapter) in order to vividly demonstrate the concept to his students.
- 9 If, while reading the last chapter, you wondered why it is that neurons in the visual system detect lines, information theory has an answer to that, too. In 1996, Bruno Olshausen and David Field applied a similar technique as Lewicki to show that lines are what you'd expect neurons to respond to if they are encoding images efficiently.