Why do we perceive logarithmically?

Why do small children place 3 halfway between 1 and 10? Why do two light bulbs not seem twice as bright as one? Why do we perceive so many things logarithmically? **Lav R. Varshney** and **John Z. Sun** can explain it, through evolution, the statistics of nature – and reducing error.

How we perceive the world

Whether it is hearing a predator approaching or a family member talking, whether it is perceiving the size of a cartload of chimpanzees or a pride of lions, whether it is seeing a ripe fruit or an explosion, the ability to sense and respond to the natural world is critical to our survival. Indeed, the world is a random place where all intensities of sensory stimuli can arise, from smallest to largest. We must register them all, from the buzz of an insect to the boom of an avalanche or an erupting volcano: our sensory perception must allow us to efficiently represent the statistical distributions of the natural world.

If kindergarten children are asked to point to the correct location for a spoken number word on a line segment labelled with endpoints 0 and 100, they place smaller numbers closer to 0 and larger numbers closer to 100, but they do not distribute numbers evenly. They give more space on the line to small numbers. The larger ones are crammed into a narrow space at the '100' end. If they are given the numbers 1 to 10, they put 3 at about halfway. In other words, they place numbers in a compressed logarithmic mapping. They perceive numbers logarithmically.

Fourth graders, however, map linearly rather than logarithmically. But if one asks an adult member of the Mundurucu – an Amazonian indigenous group with a reduced numerical lexicon and little formal education – to complete the same task, the logarithmic mapping occurs again. It seems that placing numbers evenly along a line is something that is learned, not innate. Without the intervention of critical educational experiences, the way we perceive numbers is logarithmic[1].

Number is important to human survival – you would want to know whether one lion is facing you or several. Indeed, it could be argued that perceiving numbers logarithmically rather than linearly could give an evolutionary advantage: it could be more important to know whether it is five lions facing you or three than to know if the deer herd you are chasing contains 100 animals or just 98. But in fact perceptual systems of all kinds display a nonlinear relationship between external stimulus and internal representation. If we double the force on your hand, it will feel like less than double the pressure. If we double the salinity of water, the taste will not be twice as salty. Nonlinear scalings that give greater perceptual resolution to less intense stimuli are ubiquitous across animal species and across sensory modalities: heaviness, pain, warmth, taste, loudness, pitch, brightness, distance, time delay, and colour saturation, among others, are all perceived this way. Moreover, these mappings between observable stimulus and our internal perception-space – these psychophysical scales and laws – are approximately logarithmic.

> Counting logarithmically could give an evolutionary advantage – and could optimise error-free estimation of how many lions you are facing

With multifarious organisms adapted to a variety of niches, biology is incredibly variable; so why is the same psychophysical law present in so many animals and so many sensory modes? In our recent work with Grace Wang and Vivek Goyal[2] at MIT, we have proposed an explanation based on information theory for why sensory perception is the way it is.

The principle that animals are well adapted to their environments, whether in structure or in behaviour, has been a powerful method to answer the "why" questions of biology in formal mathematical ways. For example, the principle precisely predicts the facet size of insect compound eyes by balancing resolution and diffraction for natural light; it predicts exactly when mammals shift their gait from walking to trotting to galloping so as to minimise energy cost of movement. We can apply it to the structure of the brain as well: for example, one of us has previously argued that the physical microarchitecture of synapses in the brain is optimal for memory storage capacity per unit volume[3].

The same principle applies to the way that the brain operates – the way that we receive, process and perceive information from the outside world. If we do it efficiently, we are at an obvious evolutionary advantage. Which invites the question: what architecture of information-processing within our brains would qualify as "most efficient"?

One answer might be "one that reduces error to a minimum". Which invites the further question: what sort of error needs to be thus reduced?

Here we can turn to the nineteenth-century physicist and biologist Ernst Weber. He used to ask how the structure of the nervous systems leads to a functional representation of external space. This question led him to the first general results in experimental psychophysics: that the change in stimulus intensity, ΔS, that will be just noticeable, ΔP, is a constant ratio K of the original stimulus intensity S:

$$\Delta P = K \frac{\Delta S}{S}$$

Two Mundurucu children – and how many butterflies? Credit: Stock Connection Blue/Alamy

In other words, we do not notice *absolute* changes in stimuli; we notice *relative* changes. Which leads to an answer to the question above: the error that needs to be reduced in the brains of organisms such as ourselves is not absolute error, but relative error.

By solving Weber's differential equation and broadly putting forth a philosophy that internal representations are tied to properties of the external world, Gustav Fechner developed what is now called the Weber–Fechner law[4]. It states that perceived intensity P is logarithmic to the stimulus intensity S (above a minimal threshold of perception S_0) – see box. It has become a centrepiece of psychophysics. But until now it has been an empirical law only. Theory to explain why it should be has been lacking.

Invoking our optimisation approach to biology, we can argue that organisms are well adapted to the statistics of natural external stimuli – the distributions of intensities that assail our sensory organs – and are also internally efficient for processing the information that results. This simple principle provides a formal methodology for explaining – indeed for predicting – the logarithmic laws which seem to govern so much of our perception of the world. By applying our optimisation approach to reducing the relative error in information-processing, we can produce a model which shows why the Weber–Fechner law should apply, and why we perceive so many things logarithmically.

Towards an optimality theory

If you are trying to minimise relative error, how should you go about doing it? In other words, how should we go about formulating an optimality theory for sensory perception? Information theory, the statistical theory of communication, was first formulated by Claude Shannon in the 1940s and provides a principled framework and mathematical language to study optimal informational systems, whether technological or biological.

One of the central insights from the prehistory of information theory was that any representation of continuous-valued signals must incur some noise or distortion. Perception must obey this rule as well. Thus the ear may receive sound over a range of frequencies, and will transmit that information through communication channels to the brain, where it is processed to form a "sensation". But somewhere along the way distortion – error – will have crept in. The sensation that is the eventual output from the brain must be as robust to – unaffected by – that distortion as possible. In perception, as we have seen, it is the relative error that is the critical one. We want to consider information-processing methodologies that biological information-processing systems might use to optimise expected relative error in the presence of the noise or distortion that necessarily arises.

In formulating a Bayesian model for understanding psychophysical scales we depend on a branch of information theory called *quantisation theory*. All humans quantise on a daily basis; we often round to the nearest dollar or pound at the supermarket or interpret approximate statistics in news articles. Quantisation is also

The Weber–Fechner law states that, above a minimal threshold of perception S_0, perceived intensity P is logarithmic to stimulus intensity S:

$$P = K \log \frac{S}{S_0}$$

Although not precisely true in all sensory regimes – indeed, several alternative psychophysical laws such as Stevens's power law have been proposed – the logarithmic scaling of perception has proven useful over the last century and a half. For example, dating to the heyday of the Bell Telephone System and still in use today, speech representation for digital telephony has used a logarithmic scaling called the μ-law to match technology to its human end-users. An alternative logarithmic scaling, the A-law, is used in countries other than the United States and Japan. The algorithms used in MP3 compression of music, and in JPEG compression of images, also use logarithmic scaling to appropriately adjust signals for human perception.

But why should the Weber–Fechner and related psychophysical laws exist? Previous explanations have invoked the chemistry of sensory receptors or the informational properties of individual neurons in the brain, but did not make connections to the functional requirements of sensory perception. Ernst Weber's original question has until now been unanswered.

an important aspect of fundamental research in science, where physical instruments can only measure a phenomenon to a small number of significant digits, or in engineering for applications where communicating or storing information is costly. In most applications, quantisation is thought of as a necessary evil, taking the simple form of dropping less significant digits. However, quantisation can be thought of as a more general problem of *compression*: how can I represent the real line with a finite number of points?

We have made the assumption that there is loss of information from where a stimulus is measured at the brain's periphery to when it is perceived at higher cognitive levels due to physical constraints. Our hypothesis is that efficient communication through this channel employs quantisation, and that the robustness of the system can be studied under a Bayesian framework. Discretising a continuous signal using a finite number of points leads to stability in information representation as well as robustness to noise and mismatch. We can expect our information-processing-efficient brain to display those qualities.

Logarithmic mapping within the brain minimises relative error in perception

We need a model of the brain in which a few values are communicated well instead of a model where a wide range are communicated noisily. It will be a quantisation model. Our assumptions imply that the brain can actually only distinguish a discrete set of perceptual levels. This means that a range of sensory inputs gets mapped to a single representative point within the brain. For example, even though there is a continuum of sound intensities that raindrops make when they hit the ground, our brains would only be able to distinguish a finite number of sound levels.

We make another assumption: that the quantisation at the perceptual level is uniform, meaning the spacings between elements of perception – of the perceived sound levels in our example – are the same. Figure 1 shows this schematically: a continuous range of sensations, S, is mapped by the quantiser function to an evenly spaced set of discrete perceptions P.

The mathematics of our argument is not complex; Interested readers can find it in our *Journal of Mathematical Psychology* paper[3]. It turns out that a quantised architecture in the brain that is Bayes optimal for maintaining immunity to relative errors in interpreting signals would map the output – the perceptual space – approximately logarithmically. Hence the logarithmic spacing of the intervals along the horizontal axis in Figure 1. One might expect the optimal mapping strategy would depend on the statistical distribution of the input and it does. Remarkably – most remarkably – it does not depend much. Logarithmic perception turns out to be nearly optimal no matter what the distribution of intensities of the incoming signal, as long as it is a power law. Remarkably also, it turns out that sensations corresponding to many natural phenomena have statistical distributions that do indeed obey a power law over a range of intensities that are of behavioural interest. George Zipf had first shown that many types of data in the physical and social sciences can be well modelled using this power-law statistical distribution[5]. So perception of sound, brightness, pressure and so on is optimal for relative error if the perception scale is logarithmic.

A perceptual model that uses signal compression

The quantisation model of perception is simple and intuitive, and gives a beautifully concise relationship between the psychophysical scale and stimulus intensity distribution. As we have seen, it provides scales of sensation that are optimal for error. It accounts for why we feel pressures, tastes, loudness, distance and other sensations on a logarithmic rather than an arithmetic scale. But it may not be suitable for all sensations. For example, numerosity is a perception that may not require sensory inputs that must pass through constrained communication channels. Yet, as we have seen, numbers are still perceived logarithmically. Is this phenomenon still a consequence of optimality as derived through quantisation theory? The answer turns out to be yes – as long as the information is further processed and stored in the brain for future use, as numbers often are.

The key idea is that if the input to the quantiser is a continuous random variable, then the output is a discrete random variable. Discrete random variables are easy to compress, and there have been many successful algorithms developed on top of fundamental information theory, due for example to the need to store language texts. Several studies have argued that the brain itself can also use signal compression algorithms. Applying signal compression to the output of the quantiser proposed above yields an amazing result: the optimal choice of perception space is always logarithmic, regardless of the distribution of the stimulus intensity distribution. If compression is used, the optimality of logarithmic

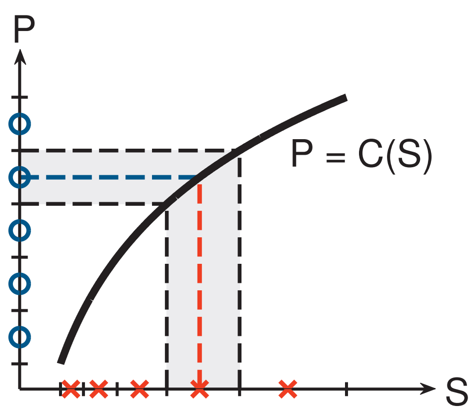

Figure 1. Quantisation model of psychophysical scales. S represents sensory stimulation, such as sound, arriving in a continuous range of intensities. Each intensity range of stimulation is mapped to a discrete point in our perception. Any sound within the shaded range of intensity is perceived as the single blue circle of loudness. The perception points are distributed linearly; the sensation ranges are distributed logarithmically

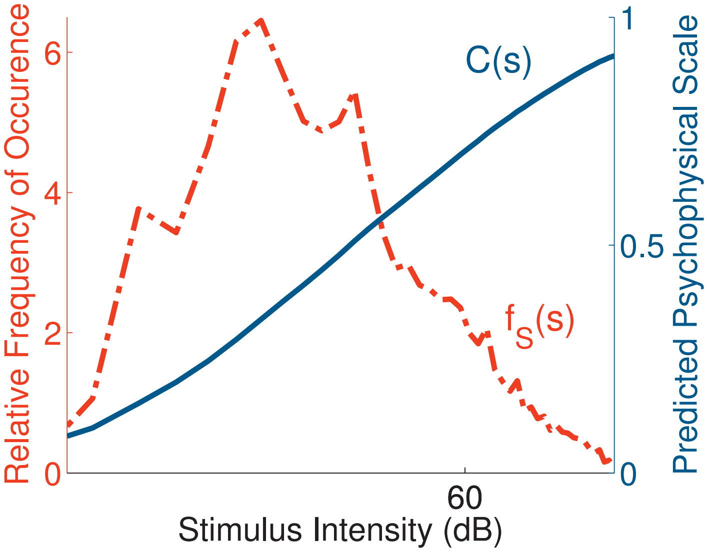

Figure 2. Predicted psychophysical scale $C(s)$ from empirical intensity distribution $f_S(s)$ of rainforest sounds

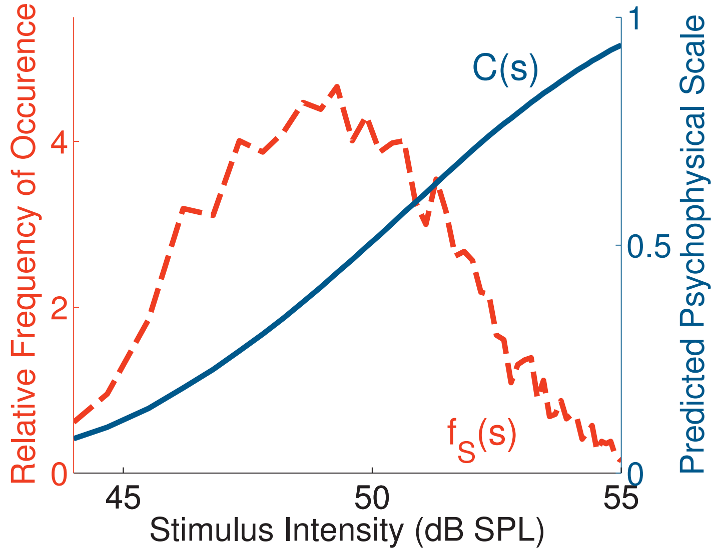

Figure 3. Predicted psychophysical scale $C(s)$ from empirical intensity distribution $f_S(s)$ of human speech

scaling laws holds for any stimulus distribution. The statistics of the input do not factor into the psychophysical scale.

A model that includes signal compression seems plausible for perception of numbers, but may be less meaningful when connected to external stimuli because properly compressing signals may require long latency and such information can be very time-sensitive.

Returning to the Amazonian jungle

We have described two models that yield psychophysical scales as Bayes optimal mappings under constraints of communication or of compression and storage. For the first model, power-law distributions of stimuli do indeed yield the Weber–Fechner logarithmic law, but we might wonder what happens when we look at actual, empirical, stimulus data. After all, data is a true test of theoretical predictions.

One well-studied sensation is the loudness of sound. Loudness is one auditory cue that conveys information related to mating rituals, predator warnings, and locations for food. It is likely that auditory systems have evolved to process natural sounds efficiently, as it is essential to survival and reproduction. Experiments have shown that the human ear can perceive loudness at differences of about 1 dB, with small deviations at the extremal ranges; this is consistent with the Weber–Fechner law.

To test our theory of optimal psychophysical scales, we return to the Amazonian rainforest of the Mundurucu: we tested our model on data sets of animal vocalisation and human speech. The animal data set features recordings of rainforest mammals from commercially available CDs; the human data set is of a male speaker reciting English sentences. In both data sets, we removed silences and then extracted sound intensity over small time windows. We then estimated the distribution of sound intensities using the empirical data. Feeding this empirical distribution of sound intensities into our quantisation model, we can predict the best psychophysical scale for minimising expected relative error.

Our results are shown in Figures 2 and 3. They demonstrate that the psychophysical scale – the perceived loudness – is approximately linear when the stimulus intensity is plotted on a log scale, which does indeed indicate that, when the expected relative error in a signal is minimised, there is an approximate logarithmic relationship between the stimulus and perception.

Conclusion

The field of psychophysics considers questions of human sensory perception, and one of its most robust findings has been the Weber–Fechner laws. It has helped us understand *how* we perceive the world. In putting forth an optimality theory for understanding *why* we perceive things as we do, we have come to a few general ideas.

Our work points out the importance of the statistics of the stimulus in interpreting psychophysical laws. Indeed, our theory allows one to predict the logarithmic scales of our perceptions simply from the statistics of the natural world. Just as Gustav Fechner had suggested in the early days, internal representations are fundamentally intertwined with the external world they describe. With an optimality result, one can say that people and animals are good processors of statistical signals – or at least our sensing apparatus seems to have evolved to be so.

Broadly speaking, our findings support the possibility that animals have information-theoretically optimal signal acquisition and perception structures at a cognitive level. Going forward, it is important to make further experimental tests of our optimality hypothesis. For example, there are certain specialised sensory modes that are known experimentally not to have logarithmic psychophysical scales. Night vision is one. Optimising for the corresponding stimulus distributions would provide a way to verify or falsify our optimal information-processing hypothesis.

References

1. Dehaene, Stanislas (2011) *The Number Sense: How the Mind Creates Mathematics*, Oxford University Press.

2. Sun, John Z., Wang, Grace I., Goyal, Vivek K. and Varshney, Lav R. (2013) A framework for Bayesian optimality of psychophysical laws, *Journal of Mathematical Psychology*, in press.

3. Varshney, Lav R., Sjostrom, P. J. and Chklovskii, D. B. (2006) Optimal information storage in noisy synapses under resource constraints, *Neuron*, **52**(3), 409–423.

4. Fechner, Gustav Theodor (1860) *Elemente der Psychophysik*, Leipzig: Breitkopf & Härtel, 1860.

5. Zipf, George K. (1949) *Human Behavior and the Principle of Least Effort*, Cambridge, MA: Addison-Wesley.

Lav R. Varshney is at the IBM T. J. Watson Research Center. John Z. Sun is in Vivek Goyal's Signal Transformation and Information Representation group at the MIT Research Laboratory of Electronics.