Review of Purves and Lotto: *"Why we see what we do: an empirical theory of vision"*

David Burr, Isituto di Neuroscienze del CNR, Pisa and Department of Psychology, University of Florence, Italy.

This book will immediately attract attention, for its ambitious title, for the scientific credentials of Dale Purves (an eminent neuro-biologist) and for the current need for a new and comprehensive theory of vision. The book is very well written in a clear didactic style and beautifully illustrated with some superb examples of visual illusions, old and new.

As the subtitle suggests, the authors advance an empirical theory of vision, suggesting that perception is determined by visual experience. Of course the idea that the visual system can be modified and calibrated by experience is hardly new, nor indeed controversial. But Purves and Lotto are claiming much more: that "experience does not merely modulate the appearance of the world" but that "the spatial perceptions elicited by geometric (or any other) stimuli are determined *entirely* by the statistical relationship between the retinal image and all its possible sources" (p144, my italics). Indeed "work over the last 50 years on the properties of visual neurones and the circuits they form has somehow been headed in the wrong direction, …. based on the suspect notion that visual circuits are in the business of detecting features, coding stimulus elements and processing them according to a set of rules" (p11).

To illustrate their theory, let us consider a specific but representative example: Mach bands, the bright and dark bands that are seen where luminance ramps meet plateaux. As the authors rightly assert "explaining these bands is clearly a necessity for any theory that purports to have rationalized perceptions of luminance" (p72). Their empirical argument goes something like this. Luminance ramps are often associated with bright and dark bands because of the nature of reflecting surfaces. For example, the luminance profile of the angular surface of a refrigerator door may vary gradually, following Lambert's law, but if the surface is specular, as many such surfaces are, it will produce clear highlights where the ramp meets the plateaux. Lowlights will also result from shading effects. Other types of ramp luminance profiles, such as cast shadows, do not have associated light and dark bands. The visual system compiles a dossier from all visual experience, from which it computes a probability density function for the range of possible percepts. Future observation of a ramp pattern will trigger a "reflex" response based on the mean of this probability density function, which will be in an attenuated version of the bright and dark bands.

I have several difficulties with this whole scheme (far too many to list here). In order for the bands to be represented in the probability density function, the system has to be capable of sensing them: it must encode the physical highlights and lowlights with photometer precision, even though they convey no useful information about object shape or texture. Obviously, if the system is capable of detecting the physical bands, it must also be able of detecting their absence, on shadows for example. Why then does it lose this capacity to discriminate the presence of bands, and group all ramp-like stimuli together? Does this mean that at some stage of development children are *more* sensitive than adults to subtle luminance changes (akin to their greater capacity to learn languages)? I know of no evidence to suggest this, and much that is against it.

Although it is not obvious from this book, Mach bands have been studied extensively and quantitatively over many years. For example, they are highly contrast-dependent, with a clear and precise threshold below which the bands disappear. The thresholds for white and dark bars are very similar (although the reasons for their physical existence on refrigerators are very different), and they vary systematically with ramp width in a predictable fashion, never occurring for ramps smaller than 4 mins arc. Importantly, contrast thresholds for Mach bands (and indeed for most stimuli) vary very little between individuals (regardless of how many refrigerators they were exposed to as babies). How can a probabilistic theory of vision predict all these facts *quantitatively* (as other theories have done) with such little variation between individuals?

What is conspicuously lacking in this book is a quantitative approach: systematic measurements of the effects of relevant variables on subject performance, combined with quantitative modelling of how their model fits the data (compared with existing models). On the few occasions that this is attempted, the measured effects are generally consistent with most off-the-shelf models of vision, and none of the manipulations addresses specifically the issue of visual experience. As such the book reads like a series of "just-so" stories, akin to the hand-wavy ideas of many evolutionary psychologists that so infuriate serious evolutionists.

I'll try to make this point more clearly with an example where evidence points to the role of experience. After the introduction of compulsory schooling in Alaska in the 1950s, the incidence of myopia increased from 5 to 80% within a single generation. Of course this does not exclude the possibility of an inheritable predisposition for myopia, nor does it elucidate the neural mechanisms linking accommodation to eye growth; but, together with other evidence, it goes a long way to supporting the more modest claim that visual experience can affect emmetropization, and that is what is blatantly lacking with this "empirical theory of perception". As their theory is fundamentally probabilistic, the authors need not mount an Indiana Jones-style expedition in search of a lost people who have never seen an upright house; sufficient to find a generation of people who have had a different *average* exposure to a particular set of stimuli, such as metallic objects with physical highlights and lowlights. If the empirical theory has any substance at all, these people *must* have higher thresholds for seeing Mach bands (which leads me to wonder how often Ernst Mach saw such objects in 19th century Czechoslovakia).

Many arguments of this book frequently suffer from what may be termed the "single-explanation problem". Does brightness contrast (or colour contrast or binocular rivalry or whatever) reflect high-level or low-level mechanisms? They repeatedly argue that as contextual manipulations can change the salience of the effect, the effects must be high-level *rather than* low-level. Is it really inconceivable that any particular effect is not due to a single mechanism, but many such mechanisms, both high and low-level? For example, does luminance gain control occur in the photoreceptors, bipolar cells or ganglion cells? Answer: "all the above" (it seems that a successful strategy tends to get used repeatedly in any biological system). There is also good evidence that binocular rivalry occurs at many levels, so the suppression gets progressively stronger as you move up the system. The evidence of contextual influences on rivalry does not exclude low-level suppression, and certainly does not require us to abandon all known theories on binocular vision. Similar arguments apply to virtually every chapter.

A particularly annoying aspect is that the book is so poorly referenced. Despite frequent citations of luminaries such as René Descarte and Bishop Berkeley, as well as the odd 17th century Jesuit friar, much work of the last few decades has either been overlooked completely, ill-cited or cited only perfunctorily. This lack of scholarship is guaranteed not only to incense everyone who has made a contribution to the field, but also causes the authors to fall into some rather obvious traps for new players.

For example, on page 168 they resurrect and replicate a little-known experiment of Sir Charles Sherrington (showing commendable scholarship when called upon). A strobe light presented to one eye becomes perceptually continuous at frequencies above about 60 Hz (the classic flicker-fusion threshold); but if flashed alternately to the two eyes, the limit remains at 60 Hz for each eye, even though the (linear) binocularly sum of the stimulus should strobe at 120 Hz. But is this surprising, given the known non-linearities of the visual system? The commonly observed *second harmonic distortion* (positive neural responses to both ups and downs of light level) will lead to virtually identical monocular ganglion cell responses, ideal for binocular summation. This (over-simplified) analysis draws no further than very basic (albeit post-Sherrington) sensory physiology, and is quite uncontroversial. Purves and Lotto, however, conclude that this experiment shows that "vision operates … without ever fusing the view of the two eyes" (p171)! A few pages later they cast severe doubts on whether "stereopsis derives from a computation of the geometric differences between the positions of the two retinas" (p179). All this is strong stuff, but with little to back it up other than the fallacious argument described above, and mentioning some interesting problems posed by stereopsis, such as Panum's limiting case and the correspondence problem. What they do not mention is that most theories of stereopsis deal adequately and elegantly with these and more recently described problems, such as Nakayama's "Da Vinci" stereopsis, at most requiring some supplementary mechanisms.

I was particularly disappointed by the chapter on spatial illusions (chapter 7), as I believe that an empirical approach should have something to offer in explaining these compelling text-book illusions (Műller-Lyer etc). However, the theory here becomes even more vague and hand-wavy, to my mind difficult to distinguish from Richard

Gregory's idea that the patterns are similar to those we have long observed on receding railway lines and the inside and outside of buildings. I remain convinced that Gregory is essentially right, although his theory probably does require some modification to deal with the objections that have been raised against it.

In the motion perception chapter they address only one relatively minor issue, the aperture problem. They show (as others have before) that the apparent direction of lines can be influenced by the angle of the aperture etc. The results follow largely the predictions of their probabilistic analysis (which is essentially an average), but also the predictions of the simpler mechanistic vector-sum model. It is curious that they chose the aperture problem to investigate, as in real life a contour-impoverished stimulus rarely moves behind an aperture. The aperture problem is interesting mainly because it applies to motion detectors themselves, given that their receptive fields have limited extent: but as far as I can understand, Purves and Lotto do not believe in receptive fields, certainly not as an intrinsic property of motion detectors. What about other well-studied phenomena of motion, such as the motion after-effect? Are we to believe this occurs because of a probabilistic tendency for opposite motion to follow extended viewing of motion in one direction? Do trains usually reverse direction when they stop? Did Robert Adaams perceive the effect on the Falls of Foyer because on all previous waterfalls the adjacent rock formation moved physically upwards whenever he shifted his gaze?

In the penultimate chapter they touch on the mechanisms by which experience could shape perception, and it will come as little surprise that Hebbian-reinforced neural networks are called into play. But again this lacks quantitative detail, with no estimates of how many observations are needed for the self-learning networks to establish, nor details of how long the critical learning period should be (or indeed if there need be one at all). This modelling is essential for the credibility of the theory, as unconstrained Baysian-type networks can very easily go off the rails. The real difficulty they are up against is the very problem that they are trying to solve: that "any element of a visual stimulus could have arisen from many – indeed *infinitely many* – different objects and conditions" (p5, my italics). It is this total lack of constraint that will make convergence so difficult unless there are many – *perhaps infinitely many* – learning trials. The treatment of the physiology is also quite superficial, failing to come up with any credible evidence in their favour and ignoring a great deal that runs contrary (such as the fact that binocular dominance columns develop almost normally during binocular deprivation, and even binocular enucleation). Whereas Marr's book of computational vision was firmly grounded on the solid and well-accepted biological facts available to him at the time, this book is strangely out of step with modern advances of neurobiology, that tend to show that experience is permissive rather than instructive.

Even before reading this book one has to ask what useful purpose is likely to be served in reviving the sterile empiricist-nativist polemic that has plagued philosophy and psychology for centuries. Surely common sense must prevail in the end. No one disputes the importance of experience for vision, for fine-tuning the system and keeping it robustly calibrated throughout life's trials and tribulations. But to suggest that our extraordinary perceptual systems are assembled entirely from probabilistic environmental associations without any "purpose-built" (evolved) hardware goes beyond the fantasy of science fiction. In any event, let us not forget that if experience is to influence the system, the system must be genetically predisposed for such adaptability, often quite a complex process (consider the myopia problem).

So despite the ambitious title, the provocative introduction, the commendable clarity of the exposition and the undisputed credentials of the first author, I am afraid that anyone with a genuine interest in *"Why we see what we do"* will be very disappointed by this volume.